1   **Temperature change has the same effect as genetic variation on a morphological trait**
2   **involved in reproductive isolation between Drosophila sister species**
3
4   Alex E. Peluffo[1,*], Mehdi Hamdani[1], Alejandra Vargas-Valderrama[1], Jean R. David[2,3],
5   François Mallard[4], François Graner[5], Virginie Courtier-Orgogozo[1]
6
7   [1]Institut Jacques Monod, CNRS, Univ. de Paris, 75013 Paris, France.
8   [2]Institut Systématique Evolution Biodiversité (ISYEB), CNRS, MNHN, Sorbonne Université,
9   EPHE, 57 rue Cuvier, CP 50, 75005 Paris, France.
10 [3]Laboratoire Evolution, Génomes, Comportement, Biodiversité (EGCE), CNRS, IRD, Univ.
11 Paris-sud, Université Paris-Saclay, 91198 Gif-sur-Yvette, France.
12 [4]Institut de Biologie de l'École Normale Supérieure, CNRS UMR 8197, Inserm U1024, PSL
13 Research University, F-75005 Paris, France.
14 [5]Matière et Systèmes Complexes, CNRS UMR 7057, Univ. de Paris, Paris, France.
15 *Current address: Data Science Department, Pharnext S.A., 11 rue René Jacques, 92130
16 Issy-les-Moulineaux, France.
17
18 Corresponding authors: VCO: virginie.courtier@normalesup.org, tel: +33 1 57 27 80 43, fax:
19 +33 1 57 27 80 87; AEP: alex.peluffo@icloud.com, tel: +33 1 41 09 22 42
20
21 **Author details**
22 AEP: alex.peluffo@icloud.com ORCID 0000-0001-7216-4205
23 MH: m_hamdani@outlook.fr   no ORCID
24 AVV: alejandra.vargas-valderrama@inserm.fr ORCID 0000-0002-6003-3459
25 JRD: Jean.David@egce.cnrs-gif.fr no ORCID
26 FM: francois.mallard@ens.fr ORCID: 0000-0003-2087-1914
27 FG: francois.graner@univ-paris-diderot.fr ORCID : 0000-0002-4766-3579
28 VCO: virginie.courtier@normalesup.org ORCID: 0000-0002-9297-9230
29
30 **Author Contributions**
31 AEP designed the study, JRD collected fly strains, MH, AVV and AEP collected the data,
32 AEP and FG designed the analysis pipeline, FM wrote the ImageJ plugin to extract contour
33 pixels, AEP, FG and VCO analyzed the data, AEP, VCO and FG wrote the paper. All authors
34 have read and approved the manuscript.
35

49 **Temperature change has the same effect as genetic variation on a morphological trait**
50 **involved in reproductive isolation between Drosophila sister species**
51
52 running title: shape plasticity to temperature & speciation
53
54 **Abstract**
55 **Phenotypic plasticity, the capacity of one genotype to generate distinct phenotypes in**
56 **different environments, is usually thought to facilitate species divergence by opening**
57 **novel ecological niches to plastic individuals. Here we reveal a case of speciation**
58 **where this "plasticity first" scenario might not hold. Male genitalia are usually**
59 **extremely divergent between closely related species, but relatively constant within**
60 **one species. Under the lock-and-key hypothesis, rapid morphological evolution is**
61 **associated with a high match between male and female genitalia of the same species**
62 **and a low match between male and females of closely related species. Previous**
63 **studies have suggested plasticity of genitalia to be a proof against the lock-and-key**
64 **hypothesis since the environmentally triggered phenotypic change could modify the**
65 **"key". Here we examine the effect of temperature on the shape of the ventral**
66 **branches, a male genital structure involved in reproductive isolation, in the sister**
67 **species *Drosophila santomea* and *D. yakuba*. We designed a semi-automatic**
68 **measurement pipeline that can reliably identify curvatures and landmarks based on**
69 **manually digitized contours of the ventral branches. With this method, we observed**
70 **that ventral branches are not plastic in *D. yakuba* but that in *D. santomea* temperature**
71 **change phenocopies interspecific genetic variation between both species for ventral**
72 **branches shape. Our results suggest that speciation of *D. santomea* and *D. yakuba***
73 **was associated with a gain of plasticity and that genitalia plasticity can be compatible**
74 **with the lock-and-key hypothesis.**
75 (234 words)
76
77 **Keywords**
78 Plasticity, Speciation, Phenotypic accommodation, Genitalia, Reproductive isolation,
79 Drosophila, Landmark, Automatic Detection, Shape analysis
80
81 **Abbreviations**
82 **SE: Standard Error**
83 QTL: Quantitative Trait Locus
84 ST: spine thrust
85
86 **Data Accessibility Statement**
87 The images, contours, scripts, and measurement values are available on DRYAD (link not
88 shown here to keep anonymity).
89
90 **Conflict of Interest**
91 The authors declare that they have no competing interests.

Introduction

92
93   Phenotypic plasticity, the capacity for one genotype to generate multiple phenotypes
94   in response to environmental variation, is a pervasive feature of biological systems (Debat &
95   David, 2001; Klingenberg, 2019). The connection between plasticity and speciation is
96   multifaceted (Lafuente & Beldade, 2019). On the one hand, plasticity can be heritable and
97   modified by selection. On the other hand, plasticity can favor adaptation and speciation. As
98   animals colonize novel habitats or face changing climate conditions, the phenotypic traits
99   that are optimal for fitness are usually different from those experienced in the ancestral
100   population. Waddington was among the first to suggest that organisms may solve this
101   challenge by phenotypic plasticity first and later on by genetic fixation of what was previously
102   an environmentally induced phenotypic trait (a process he called "genetic assimilation")
103   (Waddington, 1942). According to several authors, the trait variations enabled by plasticity
104   can initiate and accelerate the pace of adaptive evolution and promote morphological
105   diversification. This central idea is at the basis of the "flexible stem hypothesis" (West-
106   Eberhard, 2003; Schneider & Meyer, 2017), the "plasticity-first" model (Levis & Pfennig,
107   2016) or the "buying time" hypothesis, where plasticity allows the population to persist long
108   enough for adaptive mutations to arise and become fixed (Pennisi, 2018). A key feature of all
109   these views is that the phenotypic change triggered by the plastic response, which allows
110   the colonization of the new niches, is a phenocopy, i.e., that the phenotypic change can be
111   developmentally triggered by environmental variation or genetic variation interchangeably
112   (Lafuente & Beldade, 2019). As we learn more about the genes mediating phenotypic
113   plasticity (Gibert, 2017), it appears that similar phenotypic changes, either environmentally or
114   genetically induced, can sometimes involve the same genetic loci. For example, the same
115   enhancer of the gene *tan* contributes to both phenotypic plasticity in *Drosophila*
116   *melanogaster* (Gibert *et al.*, 2016) and interspecific evolution between sister-species *D.*
117   *santomea* and *D. yakuba* with respect to abdomen pigmentation (Jeong *et al.*, 2008).
118   Depending on the setting, plasticity can either accelerate, slow down, or have little
119   effect on evolution and species divergence (Price *et al.*, 2003). Speciation, the process
120   through which lineages diverge and become reproductively isolated, involves the
121   accumulation over time of barriers limiting interbreeding, including divergence in ecological
122   niches, behavioral isolation and genomic incompatibilities (Coyne & Orr, 2004). As early as
123   1844, anatomical differences in genitalia between closely related species were proposed to
124   be an essential mechanism maintaining reproductive isolation, as the so-called "lock-and-
125   key" hypothesis (Dufour, 1844; Masly, 2011). In animals with internal fertilization, genitalia
126   are the most rapidly evolving organs in terms of morphology (Eberhard, 1988), suggesting
127   that a significant part of the speciation process involves anatomical divergence in genitalia.
128   Alternatively, genital evolution can be a by-product of other evolutionary processes occurring
129   within single lineages, independently of speciation (such as sexual selection), and lead to
130   reproductive isolation as a by-product, when individuals attempt to hybridize with other
131   lineages (Masly, 2011).
132   The lock-and-key hypothesis, even in species where it seems applicable, has been
133   challenged by a variety of observations, including the facts that (1) genitalia in females do
134   not differ as much as in males, (2) closely related species with conspicuous genital
135   differences can still often produce hybrids, (3) males with laser-ablated genital organs can
136   still copulate with no observed defect, and (4) genitalia morphology can be sensitive to
137   temperature or nutrition (Shapiro & Porter, 1989; Arnqvist & Thornhill, 1998; Andrade *et al.*,
138   2005; Masly, 2011; Simmons, 2014; LeVasseur-Viens *et al.*, 2015 and references therein). It

139   is thus possible that in some taxonomic groups interspecific differences in genital
140   morphology do not contribute to reproductive isolation.
141   To better comprehend the link between plasticity and speciation, careful
142   examinations of particular cases are essential, and genital traits involved in reproductive
143   isolation represent highly relevant model systems. How plastic are genitalia in general?
144   Surprisingly, few studies have examined genitalia after raising organisms in various
145   conditions. In the water strider Aquarius remigis, the mosquito *Aedes aegypti* and the fly *D.*
146   *melanogaster*, changes in larval crowding, nutrition conditions or temperature were found to
147   affect adult body size but had little effect on the size of the external genitalia (Wheeler *et al.*,
148   1993; Fairbairn, 2005; Shingleton *et al.*, 2009). However, in two other species, the mosquito
149   *Anopheles albimanus* and the fly *Drosophila mediopunctata*, the size and shape of the male
150   intromittent organ was found to vary with rearing temperature (Hribar, 1996; Andrade *et al.*,
151   2005). Overall, analysis of individuals sampled from the wild show that for a given arthropod
152   or mammal species, the genitalia are usually more or less the same size whereas adult body
153   size varies extensively (Eberhard *et al.*, 1998; Dreyer & Shingleton, 2011 and references
154   therein). These observations are concordant with the "lock-and-key hypothesis", where male
155   genitalia have to be of a particular size and shape to physically fit with the female genitalia.
156   They are also explained by the "one-size-fits-all" hypothesis, where females appear to prefer
157   males with genitalia of intermediate size (Eberhard *et al.*, 1998).
158   In order to analyze and quantify the possible link between plasticity, reproductive
159   isolation and interspecific divergence, we chose to examine the effect of temperature on a
160   male primary sexual trait likely involved in reproductive isolation between two Drosophila
161   sister species, *D. santomea* and *D. yakuba*. These two species form an attractive system
162   because their natural environment is relatively well characterized, they are known to
163   hybridize, and one of their most remarkable morphological difference is a primary sexual trait
164   that seem to be involved in a "lock-and-key" mechanism. *D. santomea* and *D. yakuba*
165   diverged approximately 0.5-1 million years ago (Turissini & Matute, 2017). They can be
166   crossed to generate fertile F1 females (Lachaise *et al.*, 2000). *D. santomea* is endemic to the
167   island of São Tomé, a volcanic island off the coast of Gabon (Lachaise *et al.*, 2000), while *D.*
168   *yakuba* is found in São Tomé and throughout sub-Saharan Africa (Lachaise *et al.*, 1988,
169   2000). In São Tomé, *D. santomea* lives in the mist forests at high elevations while *D. yakuba*
170   is found in open habitats associated with human presence, mostly at low elevations (Llopart
171   *et al.*, 2005a; b) Both species co-occur at mid-elevation, around 1150 m, and hybrids have
172   been found consistently in this hybrid zone since its discovery in 1999 (Lachaise *et al.*, 2000;
173   Llopart *et al.*, 2005a; Comeault *et al.*, 2016; Cooper *et al.*, 2018). *D. santomea* being insular,
174   it is thought that this species originated from a common ancestor with *D. yakuba*, which
175   colonized the island about 0.5-1 million years ago (Cariou *et al.*, 2001; Llopart *et al.*, 2002;
176   Turissini & Matute, 2017), and that the present co-occurence of *D. santomea* and *D. yakuba*
177   in São Tomé reflects secondary colonization by *D. yakuba* from the African mainland, maybe
178   during the last 500 years when Portuguese colonised the island (Cariou *et al.*, 2001).
179   Analysis of genomic and mitochondrial DNA sequences indicate that gene flow occurred
180   between the *D. santomea* and *D. yakuba* more than 1,000 generations ago (Turissini &
181   Matute, 2017; Cooper *et al.*, 2019).
182   Multiple potential reproductive isolating mechanisms have been identified between
183   the two species, such as genetic incompatibilities (Coyne *et al.*, 2004; Moehring *et al.*, 2006),
184   ecological niche divergence (Matute *et al.*, 2009), mate discrimination (Lachaise *et al.*, 2000;
185   Coyne *et al.*, 2002), behavioral (Cande *et al.*, 2012), physiological (Matute, 2010) and
186   morphological differences (Lachaise *et al.*, 2000; Jeong *et al.*, 2008; Nagy *et al.*, 2018; Liu *et*

4

187   *al.*, 2019). One reproductive isolating mechanism between *D. yakuba* and *D. santomea*
188   involves a difference in ventral branches shape in the male genitalia and is the most
189   conspicuous difference in male genitalia shape between the two species (Kamimura &
190   Mitsumoto, 2012b; Yassin & Orgogozo, 2013) (Figure 1). Ventral branches are located
191   below the aedeagus [i.e., the insect phallus (Rice et al., 2019)] and are only found in the *D.*
192   *yakuba* complex, which comprises D. teissieri, D. yakuba and *D. santomea* (Yassin and
193   Orgogozo 2013).
194   In *D. yakuba*, spiny ventral branches insert inside female protective pouches during mating.
195   In *D. santomea*, the male spines and female pouches are absent. These structures appear
196   to play important roles during copulation. When mating with *D. yakuba* males, *D. santomea*
197   females are wounded by the spines of the male ventral branches and they also live shorter
198   than females mating with conspecific males (Matute & Coyne, 2010; Kamimura, 2012;
199   Kamimura & Mitsumoto, 2012b). Compared to D. teissieiri females, D. santomea females
200   also survive less to interspecific copulation with *D. mauritiana* (Yassin & David, 2016).
201   Moreover, Kamimura and Mitsumoto (2012b) reported that "copulating pairs of D. santomea
202   males x D. yakuba females dislodge readily when disturbed", suggesting that the spines may
203   fasten genital coupling (Masly, 2011). We previously found that a major QTL on
204   chromosome 3L contributes to the ventral branches shape difference between *D. santomea*
205   and *D. yakuba* (Peluffo *et al.*, 2015).
206   In São Tomé the climate is very stable throughout the year, with only a 2.5°C- difference
207   between the average daily temperature of the warmest month (March) and of the coldest
208   one (July), and daily oscillations of about 5°C only (climate-data.org,
209   www.worldclim.org/bioclim). Based on temperature measurements at Monte Café (climate-
210   data.org), we estimate that the average temperature in the hybrid zone of Bom Sucesso
211   (1153m) varies between 15.5°C and 18°C throughout the year. In the wild, *D. santomea* flies
212   are thus likely developing mainly at temperatures around 18°C or lower.
213   In previous studies of ventral branch shape, flies were raised either at 21°C (Yassin
214   & Orgogozo, 2013) or 25°C (Kamimura, 2012; Kamimura & Mitsumoto, 2012b; Peluffo *et al.*,
215   2015). Here, we report that *D. santomea* males raised at 18°C develop spiny ventral
216   branches comparable to those of *D. yakuba* raised at 25°C. This is a surprising example
217   where organs potentially directly linked with reproductive isolation undergo a plastic
218   modification similar to the difference between two sister species. To better characterize the
219   morphological change in ventral branches shape, we developed a user-friendly method to
220   quantify contour curvatures and automatically detect spines using machine learning. We
221   used it to examine the plastic response of ventral branches development at 18°C and 25°C
222   both in newly collected wild strains and in strains kept in the lab for many years.
223
224

225   **Material and Methods**
226   **Fly rearing and imaging**
227   Fly strains (Table 1) were kept at 22˚C on standard yeast-cornmeal–agar medium in
228   uncrowded conditions before the beginning of the experiments. For each strain, roughly 20
229   individuals were transferred from the 22˚C stock to either 18˚C or 25°C, kept for a minimum
230   of two non-overlapping adult generations. Adult males were 5 to 7 days old when frozen at −
231   80°C for subsequent dissection. Dissection of genitalia was performed in 1X PBS at room
232   temperature. Each genitalia was mounted on standard glass slides in DMHF (Dimethyl

233   Hydantoin Formaldehyde, Entomopraxis) medium and kept overnight before imaging on an
234   Olympus IX83 inverted station at 40X.
235
236
237

Table 1. List of Isofemale lines used in this study. For each species, the most common name in the literature, the location, year of capture and reference to origin of the strain are given. All lines are indicated in the same order as in Figure 2.

Species	Name	Location	Year	Reference
D. yakuba	Ivory Coast	Ivory Coast	1955	Cornell National Drosophila Species Stock Center, Strain #14021-0261.00 (given by D. Stern)
D. yakuba	BM2015	São Tomé, Bom Sucesso Botanical Garden, 1150m	February 2015	This study
D. yakuba	Oku	Cameroun, Mt. Oku, 2000m	April 2016	This study
D. yakuba	Raphia	Cameroun, Mt. Oku, 1800m	April 2016	This study
D. santomea	STO.4	São Tomé, Obo Natural Reserve, submontane forest 1300-1450m	1998	(Lachaise *et al.*, 2000) Cornell National Drosophila Species Stock Center, Strain #14021-0271.00 (given by D. Stern)
D. santomea	STO Cago 1482	São Tomé 1482m	2001	(Llopart *et al.*, 2005b) This strain's original name is STO-LAGO 1482 (given by D. Stern)
D. santomea	Quija22	São Tomé, Quija River, 650m	2009	(Gavin-Smyth & Matute, 2013) This strain name is also Quija650.22 (given by D. Matute)
D. santomea	BM152	São Tomé Bom Sucesso Botanical Garden, 1150m	February 2015	This study
D. santomea	BM153	São Tomé Bom Sucesso Botanical Garden, 1150m	February 2015	This study
D. santomea	BM161	São Tomé Bom Sucesso Botanical Garden, 1150m	September 2016	This study
D. santomea	BM167	São Tomé Bom Sucesso Botanical Garden, 1150m	September 2016	This study
D. santomea	1563	EYFP lab strain derived from STO CAGO 1482, Insertion @ 3L:11.843.137	2001	(Stern *et al.*, 2017) (given by D. Stern)

7

246 **Raw contour acquisition**

247   All contours were digitized by the same person. Pictures were anonymized for manual

248 contour acquisition so that the digitizer did not know the genotype. Digitization was skipped

249 when the quality of the mounting was judged to be poor. For each picture, a custom ImageJ

250 plugin was used to extract *x,y* coordinates (in pixels) of the contour. The plugin is designed

251 to open all the pictures contained in a directory, allowing the user to manually draw a contour

252 of the object of interest using the freehand tool of ImageJ.The raw contour is a series of

253 points p_1, p_2, ..., p_n in a two-dimensional space *x,y* where *n* is the number of points over

254 which the contour passes (usually $500 < n < 1000$). The contour is open and its endpoints

255 are unimportant (Figure 1). It is analyzed (and twice smoothed) as follows.

256

257 **Smoothed contour**

258 The first layer of transformation is a rectangular smoothing filter over the raw contour to

259 obtain the smoothed contour. At each point p_j with coordinates (x_j, y_j), we derive p'_j with

260 coordinates (x'_j, y'_j) where $x'_j = \frac{1}{2\alpha n} \sum_{i=j-\alpha n}^{j+\alpha n} x_i$ and $y'_j = \frac{1}{2\alpha n} \sum_{i=j-\alpha n}^{j+\alpha n} y_i$.

261 Here the contour smoothing parameter *α*, to be adjusted via learning, describes the

262 proportion of points (relative to the total number of points forming the contour) to include in

263 the smoothing. This implies that the smoothed contour is *2αn* points shorter (*αn* on each

264 side) than the raw contour.

265

266 **Raw curvature of the smoothed contour**

267 For each smoothed contour, the raw curvature *k* is computed with a sliding window of three

268 points. For any set of three points *M*, *N*, *P* forming a triangle, the diameter of the

269 circumscribed circle to this triangle, $2r = MP / \sin(MN,NP)$ can be computed as the product

270 of the Euclidean distances divided by the cross product of the two (non-basal) sides, *r* is the

271 curvature radius in *N*, and the curvature *k* in *N* is the inverse of *r* :

272

273 $$k = \frac{1}{r} = 2\frac{|\overrightarrow{MN} \times \overrightarrow{NP}|}{MN \cdot NP \cdot PM}$$

274

275 The flatter the contour, the wider the circumscribed circle, the larger the radius r, and the

276 smaller the curvature k. For each contour, the curvature profile is the curvature k_j computed

277 over p'_j in $[p_{'1}, p_m]$ using its neighbouring points (M, N, $P = p'_{j-1}, p'_j, p'_{j+1}$) versus the

278 curvilinear abscissa s_j of p'_j which is the sum of Euclidean distances from origin, $s_j = p'_1 p'_2 +$

279 $p'_2 p'_3 + \ldots + p'_{j-1} p'_j$.

280

281 **Refined curvature**

282 We then use this first raw curvature estimation as information to refine the curvature in a

283 second pass. In this second measure, the refined curvature k_j' is computed over an adaptive

284 window of size $a = \frac{1}{|k|}$ for $k < 0.1$ and $a = 10$ otherwise: M, N, $P = p'_{j-a}$, p'_j, p'_{j+a}. This means

285 that the curvature is computed over a larger distance where it is small (and curvature radius

286 is large), which requires more smoothing, without losing the sharpness of curvature peak

287 determination where the curvature is large.

288

289

290

Smoothed curvature

To improve curvature signal to noise ratio, for each point p'_j with coordinates (x'_j, y'_j) and refined curvature k'_j, we compute the smoothed curvature k''_j as a weighted moving average with triangular weights:

$$k"_j = \frac{\sum_{i=j-\beta n}^{j+\beta n} w_i \; k_i}{\sum_{i=j-\beta n}^{j+\beta n} w_i}$$

with $w_j = \beta n, \ldots, w_i = \beta n - |i - j|, \ldots, w_{j-\beta n} = w_{j+\beta n} = 0$ and where β is the smoothing parameter to adjust via learning. β describes the proportion of points (relative to the total number of contour points) to include in the smoothing. This implies that the smoothed curvature contour is $2\beta n$ points shorter (βn on each side) than the smoothed contour.

Landmark detection

Curvature around the start and end of the contour is noisy; it corresponds to a region of low curvature, at the beginning and end of the contour, outside of the region where we expect to find the five landmarks (Figure 1). In addition, the contour digitization by the user, which tends to start at a precise point and to end in a long stroke, results in a slight left-right asymmetry in the curvature profile. After superimposing all smoothed curvature profiles, we choose to exclude the first and last 20% of the smoothed contour. We find that the axis of symmetry (midline) is at position 0.475 instead of 0.5 for a symmetric profile. Landmarks are Bookstein's type 2 (local maxima of curvature) (Bookstein, 1992): maxima of the smoothed curvature for landmarks 1, 3, 5 and minima for landmarks 2 and 4. Having detected all minima and maxima, we first define landmark 3 as the maximum closest to the midline position, landmark 2 as the lowest minimum to the left of landmark 3 and landmark 1 as the maximum closest to landmark 2. Following the same logic, we define landmark 4 as the lowest minimum to the right of landmark 3 and landmark 5 as the maximum closest to landmark 4. Having detected all five landmarks, we found that there can seldom be more than one maximum between landmark 2 and landmark 4. In such situation, we allow resampling of landmark 3 to the highest maximum between landmarks 2 and 4. Finally, we exclude individuals that do not display all five landmarks after detection.

Spine thrust measure

Having detected all five landmarks, we quantify form using a measure previously introduced (Peluffo *et al.*, 2015), which is highly correlated to the Procrustes analysis principal component measure of inter-specific form variation, and which we called "spine thrust" (ST). ST is a measure of how much spines are elevated above the central ridge of the ventral branches and is computed as:

$$ST = \frac{1}{2}(Y_{L1} + Y_{L5}) - Y_{L3}$$

where Y_{L1}, Y_{L3} and Y_{L5} are the *Y* coordinate of landmarks 1, 3, 5, respectively. This measurement depends on the precise definition of *X* and *Y* axes. Here the *X*-axis is defined as the axis passing by landmarks 2 and 4 and oriented from 2 to 4, and with the *Y*-axis defined so that (X, Y) is an oriented orthonormal basis.

Machine learning

Detection of maxima and minima is a simple feature detection that relies on the derivative of the smoothed curvature profile. However, there are two parameters *α*, *β*, one for each smoothing filter (contour and curvature), which modulate the number and position of these

9

336 detected maxima and minima. It is possible to explore a set of values for α and β such that
337 the correlation between manually digitized landmarks and automatically detected landmarks
338 is optimized. Given that humans may introduce bias in the positioning of the landmarks, the
339 human output may not be optimal over the machine output. This is why we chose not to
340 quantify the learning success rate of our algorithm using the Area Under the Receiver
341 Operating Characteristic Curve but instead to search for combinations of parameter values
342 which yield the highest Pearson correlation value r^2 for ST measured over manually
343 digitized landmarks versus ST measured with automatically digitized landmarks.
344
345 **Statistical analyses**
346 All statistical analyses were performed using R version 3.4.3 (R Team 2016). We performed
347 two different sets of statistical analyses to investigate how ST changes across species, year
348 of collection and temperature. First, we fitted a standard multiple linear regression with
349 species, year of collection and temperature as numeric predictors using the standard R
350 function lm(). We chose the best model based on the variance explained provided by the
351 r^2 value. Table 2 presents the output of the lm() function using the R package jtools (v1.0.0)
352 (https://cran.r-project.org/web/packages/jtools/jtools.pdf) and its function export_summs()
353 with "scale" and "transform.response" set to "TRUE" which scales and centers the response
354 variable and reports standardized regression coefficients with their heteroskedasticity-robust
355 standard errors (see Table S1 for the raw lm() output). Second, we performed a regression
356 tree analysis and performed cross-validation using recursive partitioning with the regression
357 trees R package "rpart" version 4.1.13 (Therneau *et al.*, 2018) and the associated function
358 rpart() with the "anova" method and obtained the approximate r^2 from a 10-fold cross-
359 validation using the rsq.rpart() function. To confirm the importance of each factor on ST
360 change, we also performed random forest regression analysis using the R package
361 "RandomForest" version 4.6.14 and the randomForest() function in order. Both sets of
362 statistical analyses investigate the role of predictors in explaining a significant part of the
363 variance, multiple linear regression allows the use of interaction terms while regression trees
364 are easier to interpret (James *et al.*, 2013). In addition to these analyses, we systematically
365 plot distribution of ST across predictors (Fig. 2) showing individual values together with
366 mean, standard errors (which directly inform about two-by-two statistical significance
367 between groups), median, quartiles and estimates of the 95% confidence interval of the
368 medians, calculated as $\pm 1.58 \times \frac{IQR}{\sqrt{n}}$ where IQR is the interquartile range and n the number
369 of individuals for that IQR (Chambers *et al.*, 1983).
370
371 **Results**
372 **Spine thrust (ST) can be measured semi-automatically**
373 We previously reported that the shape of ventral branches in *D. santomea*, *D. yakuba* and
374 their hybrids can be characterized with a set of five manually detected landmarks, which
375 allows to calculate via simple arithmetic how much the lateral spines rise above the central
376 ridge, as a quantitative value named "spine thrust" (ST), expressed in micrometers (Peluffo
377 et al. 2015). The manual positioning of landmarks requires each point to be carefully
378 positioned on the exact feature for the ST measure to be exact. It can introduce between-
379 user and between-sample variability. In particular, the positioning of the three central
380 landmarks can be equivocal and may differ between users.
381   To use a less biased approach and automate the process, we decided to develop a
382 new measurement method that relies on manually digitized contours of the ventral branches,

383   which are easier to define than landmarks. The position contour of the ventral branches was
384   digitized by hand at an approximately four times faster rate than landmark detection,
385   because it can be done in a single stroke with a digital pen and the resulting ST measure is
386   barely sensitive to the exact pen position. We designed a pipeline that automatically
387   identifies the five landmarks based on the curvature of the manually digitized contours of the
388   ventral branches, and then calculates ST (Fig. 1). The typical rounded form of *D. santomea*
389   is then characterized by a null or negative value of ST (Figure 1, central panel) whereas the
390   spiny form of *D. yakuba* is characterized by a positive value of ST (Fig. 1, right panel). Note
391   that our method does not separate size and shape (Klingenberg, 2016), but considers
392   morphological form as a single quantifiable entity.
393   To assess repeatability, we digitized twice, at one month interval (at the beginning
394   and at roughly the midpoint of the digitizing effort), 30 individuals of the most characteristic
395   *D. yakuba* strain (Oku, sharp spines) and 31 individuals of the *D. santomea* strain which is
396   the most divergent from this *D. yakuba* strain (1563, extremely rounded shape and small
397   spines). We found a good correspondence   [2]between the two sets of automatic measures
398   (Fig. S2), indicating that our pipeline produces robust quantification of ventral branch form.
399
400   **Learned *α* and *β***
401   We find that the same set of 30 *D. yakuba* and 31 *D. santomea* individuals is enough to
402   identify optimal parameter values for *α* (contour smoothing) and *β* (curvature smoothing).
403   We find that with *α* = 0.025 and *β* = 0.055 we obtain $r^2 = 0.91$ (Figure S1). Although a few
404   other combinations of *α* and *β* yield the same r^2 (Figure S1), we choose this set because it
405   is the one which applies the lowest degree of smoothing.
406
407   **Strong interspecific difference in ST**
408   In total, with our semi-automated method (and after removing $n = 71$ individuals
409   incorrectly dissected or mounted, 12% of total samples, with no apparent distribution bias),
410   we phenotyped 684 individuals raised at 18°C or 25°C throughout their development,
411   corresponding to four *D. yakuba* lines and seven *D. santomea* lines collected between 1998
412   and 2016 (Table 1). We checked all the automatically detected landmarks by eyes and found
413   that 30 individuals were incorrectly digitized, with a few landmarks either missing or
414   aberrantly positioned (see Fig. S3 for a sample of such individuals), and we excluded these
415   individuals (4% of 684) from subsequent analysis. These aberrant landmark profiles were
416   found in almost all the lines and at both temperatures, with no apparent distribution bias.
417   At 18°C and 25°C, for *D. santomea* and *D. yakuba*, in all 11 wild isofemale strains,
418   we observed within-strain variability in ST values (Figure 2A, for all groups, *n* per group is
419   between 26 and 31). At both temperatures, the mean ST of each of the seven *D. santomea*
420   strains is inferior to the mean ST of any of the four *D. yakuba* strain (Figure 2A). All *D.
421   yakuba* individuals have a positive ST, while most *D. santomea* strains have a mean ST
422   close to 0 (Fig. 2A). Accordingly, multiple linear regression analysis where the best fit model
423   is *ST ~ species x years x temperature* shows that the *species* independent variable explains
424   a significant part of the variance in ST (*p*< 0.001, Table 2). Overall, and despite within strain
425   variability as well as sensitivity to temperature variation, we confirm a morphological
426   difference of ventral branches between wild strains of *D. santomea* and *D. yakuba* using our
427   semi-automatic method of form quantification based on ST (Figure 2A, Table 2).
428

429 **Table 2. Results of linear model fitting.** This best model shows the contribution of each explanatory
430 variable, considered as a numerical value,and their interactions to the overall variance of ST in the full
431 *D. santomea, D. yakuba* dataset (shown in Figures 2A-B) at both temperatures (18°C and 25°C)
432 across all years using the standard R function lm(). The standardized effect values and their
433 heteroskedasticity-robust standard errors are reported together with the range of their *p*-values. For
434 example, species having an overall effect of 1.67 implies that going from *D. santomea* to *D. yakuba*
435 (all other things being equal) increases spine thrust absolute value by a relative (compared to the
436 other effects), dimension-less, value of 1.67. The raw effects together with the full output of the model
437 are provided in Table S1 (see methods). SE: standard error.

438

Factor $n = 584$, $r^2 = 0.76$	Effect	SE	*p*-value	Significance
Species	1.67	0.06	< 0.001	***
Years	0.54	0.06	< 0.001	***
Temperature	-0.29	0.05	< 0.001	***
Species x Years	-0.56	0.07	< 0.001	***
Species x Temperature	0.31	0.09	< 0.001	***
Years x Temperature	-0.23	0.09	< 0.05	*

439
440

Ventral branches of *D. santomea* are plastic to temperature

441
442   For *D. santomea*, in all strains but the oldest one collected in 1998, the mean ST is
443   systematically smaller at 25°C compared to 18°C and standard errors do not overlap (Fig. 2).
444   In contrast, no significant difference in mean ST between 25°C and 18°C is observed for *D.*
445   *yakuba* strains, except for one strain collected in 2016 (*D. yakuba* Raphia) (Figure 2B).
446   Multiple linear regression analysis supports a negative effect of temperature, as seen with *D.*
447   *santomea* (p< 0.001, Table 2) and that effect is dependent on species (p< 0.001, Table 2).
448   For the most recently collected wild strain of *D. santomea* (BM16.2), we compared the
449   contours of the two most representative individuals raised at 18°C and 25°C, i.e., the two
450   individuals with ST values closest to the median value of their group. We observed that the
451   individual raised at 18°C has a more *D. yakuba*-like shape of ventral branches compared to
452   the individual raised at 25°C (Fig. 3).
453   We find that the statistically significant effect of temperature on *D. santomea* only is
454   also statistically dependent on the year at which the strain was collected (p < 0.05, Table 2).
455   In order to interpret better our statistical analysis with multiple regression, we performed a
456   10-fold cross-validated regression tree analysis on the full data set (2 species, 11 strains,
457   584 individuals). The 10-fold cross-validated error rate is 0.3% and using an additive model
458   of the shape *ST ~ species + years + temperature*. We found that the variance in the data set
459   is first best partitioned by species and that temperature partitions the data set best for strains
460   collected in 2015 and 2016 (Fig. 4, total variance explained as assessed by cross-validation
461   r^2 is 0.77). To confirm those results, we also performed a random forest regression analysis
462   with the same model as for the regression tree and found that the overall variance explained
463   is $r^2 = 0.74$ and that the rank of importance of each independent variable is species > years
464   > temperature. Altogether, our results show that in *D. santomea*, but not in *D. yakuba*,
465   ventral branches are sensitive to temperature during development and that this effect is
466   stronger in recently collected strains.
467

Temperature change phenocopies natural genetic variation for form

469   To compare the effects of temperature and of interspecific genetic variation on
470   ventral branch form, we used our previous QTL mapping dataset of ventral branch form
471   between *D. santomea* and *D. yakuba*, which comprises 365 *D. santomea* backcross
472   individuals (Peluffo *et al.*, 2015). In this previous study, all flies were reared at 25°C as we
473   found that this temperature was optimal to rear both species. The five landmarks were
474   placed manually on images of the ventral branches. A generalized Procrustes analysis was
475   performed on a set of 365 backcross progeny individuals and a larger dataset including the
476   backcross progeny, F1 hybrids and parents. We found that, in both cases, the principal
477   component PC1 explains an important part of the variance (58% in the full dataset and 41%
478   in the backcross), that they are highly correlated (r^2 = 0.996) and that PC1 in the backcross
479   is highly correlated to ST (spine thrust) (r^2 = 0.841) and not to centroid size (r^2 = 0.038).
480   This QTL mapping study revealed that a 2.7Mb locus on chromosome 3L explains
481   30% of the mean species difference in ST, meaning that replacing one *D. santomea* allele at
482   this locus with a *D. yakuba* allele leads to an increase in ST of about 3 μm (Peluffo *et al.*,
483   2015). Pooling all the *D. santomea* lines examined in the present study, we find that a
484   change in the raising temperature from 18°C to 25°C leads to an increase in ST of about 3.4
485   μm (Fig. 5). We conclude that the effect of temperature is as high as the effect of genetic
486   variation at the major interspecific genetic locus.
487   While the ST density distribution of all *D. yakuba* individuals shows little overlap with
488   the ST density distribution of all *D. santomea* individuals reared at 25°C, it overlaps more

13

489   with *D. santomea* individuals reared at 18°C (Fig. 5). Overall, our results show that
490   decreasing the temperature from 25°C to 18°C yield *D. santomea* males with spinier, *D.*
491   *yakuba*-like, ventral branches in the same way as introgressing a *D. yakuba* alleles in place
492   of a *D. santomea* allele at the major interspecific locus.
493
494
495   **Discussion**
496   **A dataset-independent simultaneous quantification of shape and size**
497   Our semi-automatic method, which relies on two simple layers of contour
498   transformation adjusted by regression based learning, is fast and allows the measure of form
499   variation through the simple outlining of ventral branches on 2D pictures. We note that in the
500   future, progress in edge detection algorithms (which for now introduce too much error to
501   measure with precision variations of the order of a few micrometers) might allow full
502   automation from pictures to form quantification.
503   Having drawn contours, we could also have relied on Fourier based analyses.
504   However, such methods require closed contours which in our case are difficult to draw since
505   the base of the ventral branches is a complex structure which cannot be easily delimited
506   from the cuticle of the ventral branches (Fig. 1). In addition, our method is more suitable for
507   contours in which very large and very small curvatures coexist. Furthermore, an important
508   limitation of morphometrics analyses on landmark data (e.g. Procrustes principal component
509   analysis) is that the PC values are dimensionless (Klingenberg, 2010) and may be difficult to
510   relate to physical features. With our simple measure of ST obtained from the automatically
511   detected landmarks, we are able to quantify and compare forms across studies. Importantly,
512   because we deal with absolute geometric measurements, our method simultaneously
513   analyzes shape and size, unlike most morphometric approaches (Claude, 2008;
514   Klingenberg, 2010). We believe this to be a strength in our case since both shape and size
515   of ventral branches probably contribute to the lock-and-key mechanism; e.g. spiny but short
516   *D. yakuba* ventral branches may not harm *D. santomea* females (Kamimura, 2012;
517   Kamimura & Mitsumoto, 2012b).
518
519   **Effect of temperature on size and shape**
520   In most insects and other ectotherms, adult body size typically increase with lower
521   temperatures (Angilletta Jr *et al.*, 2004). Bergmann's rule, which posits an increasing body
522   size with higher altitude, has been observed within the São Tomé island for the terrestrial
523   caecilian *Schistometopum thomense*, over a temperature range of 9°C (Measey & Van
524   Dongen, 2006). In contrast to other body parts, the genitalia of insects, and of *D.*
525   *melanogaster* in particular, have been reported as not, or little, plastic in response to
526   temperature or other types of environmental variation (Wheeler *et al.*, 1993; Eberhard, 2009;
527   Shingleton *et al.*, 2009). We find here that this is true also for *D. yakuba* but not for *D.*
528   *santomea*: changing the rearing temperature from 25°C to 18°C leads to an increase in
529   spine thrust in *D. santomea* male genitalia that phenocopies the morphological difference
530   observed between *D. santomea* and *D. yakuba*.
531   For each species, we find that strains raised in the same conditions display different
532   averages in ST, showing that the ventral branches form is influenced by genetic factors and
533   is able to evolve.
534   Plasticity of ventral branches form was detected for all the tested *D. santomea* strains
535   except the one that was maintained for the longest time in the laboratory. Furthermore, the
536   strains collected recently (in 2009, 2015 and 2016) display more pointed ventral branches at

537   18°C than the ones collected earlier. This suggests that as flies adapt to the laboratory
538   environment, the plasticity of ventral branches form towards temperature tends to be lost
539   and ventral branches tend to be more rounded. Recent studies show that *Drosophila* flies
540   can adapt to a laboratory environment in 20 generations only, which corresponds to about 8
541   months (Langmüller & Schlötterer, 2019).
542   Based on our experiments, we cannot fully rule out plasticity in *D. yakuba*. It is
543   possible that their genital morphology would be altered in external conditions outside of the
544   specific ones that we assayed here. In any case, we find that in our experimental conditions
545   the plasticity of genital form with respect to temperature is higher in *D. santomea* than *D.*
546   *yakuba.*
547
Laboratory observations should be complemented by analysis of wild-caught flies
549   Tests in the laboratory show that *D. santomea* flies appear to be poorly adapted to
550   high temperatures (Matute *et al.*, 2009). The optimal temperature for larval survival is 21°C
551   for *D. santomea* and 24°C for *D. yakuba.* Furthermore, when adult flies initially raised at
552   24°C are allowed to distribute themselves along a thermal gradient, they show a preference
553   for 23°C for *D. santomea* and between 26°C and 27°C for *D. yakuba* (Matute *et al.*, 2009).
554   These observations are in agreement with *D. santomea* being collected at higher altitudes
555   than *D. yakuba* in Sao Tomé. However, the reasons why the exact preferred temperature
556   values observed in the laboratory are different from the temperature values measured in the
557   geographic areas of the two species are unknown. Fly collections in Sao Tomé have mostly
558   been done on the north slopes of the island and in these areas *D. santomea* flies are found
559   at an altitude of 1150m or above (Lachaise *et al.*, 2000), which corresponds to temperatures
560   around 18°C or below (climate-data.org, www.worldclim.org/bioclim). However, we note that
561   on the southern slopes of the island a few *D. santomea* flies have also been collected at
562   lower altitudes (650m) in the dense mist forest near Rio Queijo (Matute & Coyne, 2010;
563   Nagy *et al.*, 2018). This suggests that *D. santomea* flies can also inhabit warmer regions of
564   the island and that they might be found across the native forest of Sao Tomé, which goes
565   down to sea level on the western slope of the island (Bell & Irian, 2019). Interestingly, this
566   type of coexistence is not unique on the island: two sister species of frogs closely match the
567   distribution of *D. santomea* and *D. yakuba*, respectively, with the endemic species
568   *Hyperolius thomensis* tied to wet forest habitats while its sister species *H. molleri* is in dry,
569   human-disturbed areas, and *H. thomensis* frogs have also been found in the southern forest
570   at 150 m (Bell & Irian, 2019).
571   It would be interesting to examine the genitalia of wild caught individual males of *D.*
572   *santomea* to check the form of their ventral branches at various altitudes. One possibility is
573   that at low altitudes in the southern part of the island *D. santomea* flies display rounded
574   ventral branches while in the hybrid zone with *D. yakuba* at 1150 m and at higher altitude,
575   where temperatures are 18°C or below, they have spinier ventral branches. Of note, *D.*
576   *santomea* flies have always been collected from traps and have never been observed
577   directly in their native environment: it is possible that they live in microenvironments whose
578   temperature is distinct from the one measured by climate stations (Feder *et al.*, 2000;
579   Negoua *et al.*, 2019).
580
Evolution of the plasticity of ventral branches form
582   To understand the relevance of this temperature sensitivity of genital form for the
583   past and present evolution of *D. santomea* and *D. yakuba*, more needs to be learnt about
584   their ecology and the plasticity of the ventral branches form of their closely related species,

15

585   *D. teissieri*. Ventral branches are only found in the three species of the *D. yakuba* complex,
586   *D. santomea*, *D. yakuba* and *D. teissieri* (Yassin & Orgogozo, 2013). Since ventral branch
587   form plasticity has not been studied in *D. teissieri*, it is unclear whether this plasticity to
588   temperature is an ancestral trait which has been lost in *D. yakuba* or if it is a novel trait which
589   evolved in *D. santomea* only. The species *D. teissieri* is not found in São Tomé but on the
590   mainland and a few islands of the African continent; it can hybridize with *D. yakuba* (Turissini
591   & Matute, 2017; Cooper *et al.*, 2018). In *D. teissieri* males, the spines are very long and no
592   layer of cuticle is present between them (Kamimura & Mitsumoto, 2012a; Yassin &
593   Orgogozo, 2013). In any case, even if the extent of ventral branch form plasticity in *D.*
594   *teissieri* was known, it would still be difficult to reconstruct ancestral trait states based on
595   only three species.
596   The female protective pouches, into which the spiny ventral branches of *D. yakuba*
597   males fit during copulation, were observed in *D. yakuba* but not in *D. santomea* females
598   raised at 21°C and 25°C (Kamimura & Mitsumoto, 2012b; Yassin & Orgogozo, 2013). It
599   would be interesting to check whether such pouches form in *D. santomea* females raised at
600   18°C, coinciding with the emergence of spiny ventral branches in males. Furthermore,
601   whether more pointed ventral branches in *D. santomea* males due to lower temperatures
602   affects copulation, reproduction and female physiology after mating is unknown.
603   If we assume that the São Tomé island species *D. santomea* arose from a *D.*
604   *yakuba*-like ancestor living on the African continent, one can hypothesize that regression in
605   ventral branch size and their plasticity evolved recently in the lineage leading to *D.*
606   *santomea*. Such a scenario is opposite to the most common view that posits that
607   morphological diversification tends to proceed through losses of plasticity, rather than gains
608   of plasticity ["flexible stem hypothesis" (West-Eberhard, 2003; Schneider & Meyer, 2017),
609   "plasticity-first" model (Levis & Pfennig, 2016) and "buying time" hypothesis (Pennisi, 2018)].
610   It is possible that the decrease in spine thrust that occurred during evolution in the lineage
611   leading to *D. santomea* was accompanied by a gain of ventral branches form plasticity
612   towards temperature. It is unclear whether the plasticity of ventral branches form to
613   temperature is adaptive. More knowledge about the ecology of *D. santomea* and its sister
614   species will be required to elaborate a convincing scenario to interpret the role of the ventral
615   branch form plasticity that we discovered.
616
617   **Conclusion**
618   Our data show that genitalia can be plastic to temperature and that this plasticity can evolve
619   coincidentally with speciation. Whereas the sensitivity of insect genitalia shape to
620   temperature or nutrition has been used previously as a proof against the lock-and-key
621   hypothesis (Arnqvist & Thornhill, 1998; Andrade *et al.*, 2005), our work suggests that
622   genitalia can be plastic without rejecting the lock-and-key hypothesis if the environmentally
623   induced changes do not hamper reproduction within each sister species lineage.
624
625
626
627
628   **Figures**
629
630   **Figure 1. Landmark detection for ventral branches at 18°C and 25°C**. For each
631   individual, a picture of the ventral branches is taken (top panel). The contour is digitized by
632   hand and smoothed (middle panel). The curvature along the contour is obtained by finite

633 differences, which are iterated for refining, and smoothed too (bottom panel). Smoothed
634 curvature (vertical axis), measured in inverse micrometers, is plotted along the contour,
635 starting from the leftmost point. The horizontal axis is the distance along the contour, called
636 the curvilinear abscissa, measured in micrometers. Here both axes are normalized by size
637 and represented in arbitrary units. In each plot the left dashed vertical line is the automatic
638 detection lower bound, the middle dashed line is the imputed global midpoint, the right
639 dashed line is the automatic detection upper bound (see methods). Red points represent
640 peaks and therefore curvature maxima whereas blue points represent cavities and therefore
641 curvature minima. Since *D. yakuba* is not sensitive to temperature, we only show one
642 characteristic shape. To understand where this genital structure is positioned within the male
643 genitalia, see Fig. 1 of Kamimura & Mitsumoto, 2012b.

645 **Figure 2. Ventral branches form is sensitive to temperature in *D. santomea*.** Isofemale
646 lines arranged by year of collection (see Table 1). (A) For each line, individual values (grey
647 points), median (thick black line), quartiles (colored boxplots), mean (colored points),
648 standard errors (black vertical segment over the colored points) and 95% confidence interval
649 estimates of the median (top and bottom notches) of automatically measured spine thrust
650 are shown. Each line was reared at 18°C (blue) or 25°C (red). (B) For each line, the same
651 mean and standard errors as in panel A are shown, together with the effect slope and
652 corresponding value of that effect (in µm).

654 **Figure 3. Difference in contour shape at 18°C versus 25°C within the same *D.***
655 ***santomea* isofemale strain collected in 2016**. Contours of the two individuals that have
656 the closest spine thrust value to the median value for D. santomea BM16.2 (right most strain
657 on Figure 2A) at 18°C and 25°C.

659 **Figure 4. Regression tree for spine thrust measures of all *D. santomea* and *D. yakuba***
660 **isofemale strains at both 18 and 25°C**. Each node gives the spine thrust mean of all
661 samples included in that node and the proportion of the total dataset included in that node.
662 Below each node are two alternatives: to the left the condition is true and to the right the
663 condition is false. Note that the split between *D. santomea* and *D. yakuba* happens at the
664 top, thereby suggesting that neither temperature nor years have an effect on spine thrust
665 within *D. yakuba*.

667 **Figure 5. Density distribution of spine thrust values for *D. santomea* lines at 18 and**
668 **25°C and *D. yakuba* lines**. Density distribution inferred from *D. santomea* males raised at
669 18°C (blue) and 25°C (red), *D. yakuba* males at both temperatures (orange). Distributions
670 are inferred from the total data shown in Figure 2A.

675 **Supplementary Figures**

677 **Figure S1. Parameter adjustment for the machine detection algorithm**. Training the
678 algorithm relies on two layers of transformation that are each dependent on one parameter:
679 contour coordinate smoothing (horizontal axis) and curvature profile smoothing (vertical
680 axis). Training was performed using a set of 61 individuals, 31 *D. santomea* 1563 and 30 *D.*

681 *yakuba* Oku (see Table 1) for which we manually digitized both landmarks and contours. For
682 each value of the two smoothing parameters, we performed linear regression of spine thrust
683 from manually digitized landmarks against spine thrust derived from automatically digitized
684 landmarks. The colors and values represent the r^2 from that regression. The value used for
685 all detections is contoured in white.
686
687 **Figure S2. Correlation between two sets of automatic measures from the same**
688 **dataset**. For the training dataset (31 *D. santomea* 1563 and 30 *D. yakuba* Oku), the same
689 user digitized the same contours twice at one month interval and spine thrust was
690 automatically measured. Each point represents one individual. The *y = x* (black dashed line)
691 and linear regression (full red line) are shown.
692
693 **Figure S3. Representative samples of landmarks incorrectly identified with the**
694 **machine detection algorithm**. For each example, we show the smoothed contour, the
695 corresponding curvature profile and identified landmarks.
696
697 **Table S1. Raw output of R lm (formula = ST ~ Species * Year * Temperature).**
698
699
700
701
702 **References**

703 Andrade, C.A., Hatadani, L.M. & Klaczko, L.B. 2005. Phenotypic plasticity of the aedeagus
704   of Drosophila mediopunctata: effect of the temperature. *J. Therm. Biol.* **30**: 518–523.
705 Angilletta Jr, M.J., Steury, T.D. & Sears, M.W. 2004. Temperature, growth rate, and body
706   size in ectotherms: fitting pieces of a life-history puzzle. *Integr. Comp. Biol.* **44**: 498–
707   509.
708 Arnqvist, G. & Thornhill, R. 1998. Evolution of animal genitalia: patterns of phenotypic and
709   genotypic variation and condition dependence of genital and non-genital morphology
710   in water strider (Heteroptera: Gerridae: Insecta). *Genet. Res.* **71**: 193–212.
711 Bell, R.C. & Irian, C.G. 2019. Phenotypic and genetic divergence in reed frogs across a
712   mosaic hybrid zone on São Tomé Island. *Biol. J. Linn. Soc.* **128**: 672–680.
713 Bookstein, F.L. 1992. *Morphometric tools for landmark data: Geometry and Biology*.
714   Cambridge University Press, Cambridge, UK.
715 Cande, J., Andolfatto, P., Prud'homme, B., Stern, D.L. & Gompel, N. 2012. Evolution of
716   Multiple Additive Loci Caused Divergence between Drosophila yakuba and D.
717   santomea in Wing Rowing during Male Courtship. *PloS One* **7**: e43888.
718 Cariou, M.L., Silvain, J.F., Daubin, V., Da Lage, J.L. & Lachaise, D. 2001. Divergence
719   between Drosophila santomea and allopatric or sympatric populations of D. yakuba
720   using paralogous amylase genes and migration scenarios along the Cameroon
721   volcanic line. *Mol. Ecol.* **10**: 649–660.
722 Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. 1983. *Graphical methods for*
723   *data analysis.* Wadsworth & Brooks, Boston, USA.
724 Claude, J. 2008. *Morphometrics with R*. Springer Science & Business Media.
725 Comeault, A.A., Venkat, A. & Matute, D.R. 2016. Correlated evolution of male and female
726   reproductive traits drive a cascading effect of reinforcement in Drosophila yakuba. In:
727   *Proc. R. Soc. B*, p. 20160730. The Royal Society.
728 Cooper, B.S., Sedghifar, A., Nash, W.T., Comeault, A.A. & Matute, D.R. 2018. A
729   maladaptive combination of traits contributes to the maintenance of a Drosophila
730   hybrid zone. *Curr. Biol.* **28**: 2940–2947.
731 Cooper, B.S., Vanderpool, D., Conner, W.R., Matute, D.R. & Turelli, M. 2019. Wolbachia
732   acquisition by Drosophila yakuba-clade hosts and transfer of incompatibility loci

733   between distantly related Wolbachia. *Genetics* **212**: 1399–1419.

734 Coyne, J.A., Elwyn, S., Kim, S.Y. & Llopart, A. 2004. Genetic studies of two sister species in
735   the Drosophila melanogaster subgroup, D. yakuba and D. santomea. *Genet. Res.* **84**:
736   11–26.

737 Coyne, J.A., Kim, S.Y., Chang, A.S., Lachaise, D. & Elwyn, S. 2002. Sexual isolation
738   between two sibling species with overlapping ranges: Drosophila santomea and
739   Drosophila yakuba. *Evolution* **56**: 2424–2434.

740 Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Sunderland, Massachusets.

741 Debat, V. & David, P. 2001. Mapping phenotypes: canalization, plasticity and developmental
742   stability. *Trends Ecol. Evol.* **16**: 555–561.

743 Dreyer, A.P. & Shingleton, A.W. 2011. The effect of genetic and environmental variation on
744   genital size in male Drosophila: canalized but developmentally unstable. *PLoS One*
745   **6**: e28278.

746 Dufour, L. 1844. Anatomie générale des diptères. *Ann. Sci. Nat.* **1**: 244–264.

747 Eberhard, W.G. 1988. *Sexual Selection and Animal Genitalia*. Harvard University Press.

748 Eberhard, W.G. 2009. Static allometry and animal genitalia. *Evol. Int. J. Org. Evol.* **63**: 48–
749   66.

750 Eberhard, W.G., Huber, B.A., Briceño, R.D., Salas, I. & Rodriguez, V. 1998. One size fits all?
751   Relationships between the size and degree of variation in genitalia and other body
752   parts in twenty species of insects and spiders. *Evolution* **52**: 415–431.

753 Fairbairn, D.J. 2005. Allometry for sexual size dimorphism: testing two hypotheses for
754   Rensch's rule in the water strider Aquarius remigis. *Am. Nat.* **166**: S69–S84. The
755   University of Chicago Press.

756 Feder, M.E., Roberts, S.P. & Bordelon, A.C. 2000. Molecular thermal telemetry of free-
757   ranging adult Drosophila melanogaster. *Oecologia* **123**: 460–465.

758 Gavin-Smyth, J. & Matute, D.R. 2013. Embryonic lethality leads to hybrid male inviability in
759   hybrids between Drosophila melanogaster and D. santomea. *Ecol. Evol.* **3**: 1580–
760   1589.

761 Gibert, J.-M. 2017. The flexible stem hypothesis: evidence from genetic data. *Dev. Genes*
762   *Evol.* **227**: 297–307.

763 Gibert, J.-M., Mouchel-Vielh, E., De Castro, S. & Peronnet, F. 2016. Phenotypic plasticity
764   through transcriptional regulation of the evolutionary hotspot gene tan in Drosophila
765   melanogaster. *PLoS Genet.* **12**: e1006218.

766 Hribar, L.J. 1996. Larval rearing temperature affects morphology of Anopheles albimanus
767   (Diptera: Culicidae) male genitalia. *J. Am. Mosq. Control Assoc.-Mosq. News* **12**:
768   295–297.

769 James, G., Witten, D., Hastie, T. & Tibshirani, R. 2013. Tree-based methods. In: *An*
770   *introduction to statistical learning*, pp. 303–335. Springer.

771 Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J. & Carroll, S.B. 2008. The evolution
772   of gene regulation underlies a morphological difference between two Drosophila
773   sister species. *Cell* **132**: 783–793.

774 Kamimura, Y. 2012. Correlated evolutionary changes in Drosophila female genitalia reduce
775   the possible infection risk caused by male copulatory wounding. *Behav. Ecol.*
776   *Sociobiol.* **66**: 1107–1114.

777 Kamimura, Y. & Mitsumoto, H. 2012a. Genital coupling and copulatory wounding in
778   Drosophila teissieri (Diptera: Drosophilidae). *Can. J. Zool.* **90**: 1437–1440.

779 Kamimura, Y. & Mitsumoto, H. 2012b. Lock-and-key structural isolation between sibling
780   Drosophila species. *Entomol. Sci.* **15**: 197–201.

781 Klingenberg, C.P. 2010. Evolution and development of shape: integrating quantitative
782   approaches. *Nat. Rev. Genet.* **11**: 623.

783 Klingenberg, C.P. 2019. Phenotypic plasticity, developmental instability and robustness: the
784   concepts and how they are connected. *Front. Ecol. Evol.* **7**: 56.

785 Klingenberg, C.P. 2016. Size, shape, and form: concepts of allometry in geometric
786   morphometrics. *Dev. Genes Evol.* **226**: 113–137. Springer.

787 Lachaise, D., Cariou, M.-L., David, J.R., Lemeunier, F., Tsacas, L. & Ashburner, M. 1988.

788   Historical biogeography of the Drosophila melanogaster species subgroup. In:
789   *Evolutionary biology*, pp. 159–225. Springer.
790   Lachaise, D., Harry, M., Solignac, M., Lemeunier, F., Bénassi, V. & Cariou, M.L. 2000.
791   Evolutionary novelties in islands: Drosophila santomea, a new melanogaster sister
792   species from São Tomé. *Proc. Biol. Sci.* **267**: 1487–1495.
793   Lafuente, E. & Beldade, P. 2019. The genomics of developmental plasticity: recent progress
794   in animal models. *Front. Genet.* **10**: 720.
795   Langmüller, A.M. & Schlötterer, C. 2019. Low concordance of short-term and long-term
796   selection responses in experimental Drosophila populations. *bioRxiv* 759704.
797   LeVasseur-Viens, H., Polak, M. & Moehring, A.J. 2015. No evidence for external genital
798   morphology affecting cryptic female choice and reproductive isolation in Drosophila.
799   *Evolution* **69**: 1797–1807. Wiley Online Library.
800   Levis, N.A. & Pfennig, D.W. 2016. Evaluating 'plasticity-first'evolution in nature: key criteria
801   and empirical approaches. *Trends Ecol. Evol.* **31**: 563–574.
802   Liu, Y., Ramos-Womack, M., Han, C., Reilly, P., Brackett, K.L., Rogers, W., *et al.* 2019.
803   Changes throughout a Genetic Network Mask the Contribution of Hox Gene
804   Evolution. *Curr. Biol.*
805   Llopart, A., Elwyn, S., Lachaise, D. & Coyne, J.A. 2002. Genetics of a difference in
806   pigmentation between Drosophila yakuba and Drosophila santomea. *Evolution* **56**:
807   2262–2277.
808   Llopart, A., Lachaise, D. & Coyne, J.A. 2005a. An anomalous hybrid zone in Drosophila.
809   *Evol. Int. J. Org. Evol.* **59**: 2602–2607.
810   Llopart, A., Lachaise, D. & Coyne, J.A. 2005b. Multilocus analysis of introgression between
811   two sympatric sister species of Drosophila: Drosophila yakuba and D. santomea.
812   *Genetics* **171**: 197–210.
813   Masly, J.P. 2011. 170 years of "lock-and-key": genital morphology and reproductive isolation.
814   *Int. J. Evol. Biol.* **2012**.
815   Matute, D.R. 2010. Reinforcement of gametic isolation in Drosophila. *PLoS Biol.* **8**:
816   e1000341.
817   Matute, D.R. & Coyne, J.A. 2010. Intrinsic reproductive isolation between two sister species
818   of Drosophila. *Evol. Int. J. Org. Evol.* **64**: 903–920.
819   Matute, D.R., Novak, C.J. & Coyne, J.A. 2009. Temperature-based extrinsic reproductive
820   isolation in two species of Drosophila. *Evol. Int. J. Org. Evol.* **63**: 595–612.
821   Measey, G.J. & Van Dongen, S. 2006. Bergmann's rule and the terrestrial caecilian
822   Schistometopum thomense (Amphibia: Gymnophiona: Caeciliidae). *Evol. Ecol. Res.*
823   **8**: 1049–1059.
824   Moehring, A.J., Llopart, A., Elwyn, S., Coyne, J.A. & Mackay, T.F. 2006. The genetic basis of
825   postzygotic reproductive isolation between Drosophila santomea and D. yakuba due
826   to hybrid male sterility. *Genetics* **173**: 225–233.
827   Nagy, O., Nuez, I., Savisaar, R., Peluffo, A.E., Yassin, A., Lang, M., *et al.* 2018. Correlated
828   evolution of two copulatory organs via a single cis-regulatory nucleotide change.
829   *Curr. Biol.* **28**: 3450–3457.
830   Negoua, H., Chakir, M., David, J.R. & Capy, P. 2019. Climatic adaptation in Drosophila:
831   phenotypic plasticity of morphological traits along a seasonal cycle. In: *Annales de la*
832   *Société entomologique de France (NS)*, pp. 48–60. Taylor & Francis.
833   Peluffo, A.E., Nuez, I., Debat, V., Savisaar, R., Stern, D.L. & Orgogozo, V. 2015. A Major
834   Locus Controls a Genital Shape Difference Involved in Reproductive Isolation
835   Between Drosophila yakuba and Drosophila santomea. *G3 Bethesda Md*, doi:
836   10.1534/g3.115.023481.
837   Pennisi, E. 2018. *Buying time*. American Association for the Advancement of Science.
838   Price, T.D., Qvarnström, A. & Irwin, D.E. 2003. The role of phenotypic plasticity in driving
839   genetic evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **270**: 1433–1440.
840   Rice, G., David, J.R., Kamimura, Y., Masly, J.P., Mcgregor, A.P., Nagy, O., *et al.* 2019. A
841   standardized nomenclature and atlas of the male terminalia of Drosophila
842   melanogaster. *Fly (Austin)* **13**: 51–64.

Schneider, R.F. & Meyer, A. 2017. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Mol. Ecol.* **26**: 330–350.

Shapiro, A.M. & Porter, A.H. 1989. The lock-and-key hypothesis: evolutionary and biosystematic interpretation of insect genitalia. *Annu. Rev. Entomol.* **34**: 231–245. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

Shingleton, A.W., Estep, C.M., Driscoll, M.V. & Dworkin, I. 2009. Many ways to be small: different environmental regulators of size generate distinct scaling relationships in Drosophila melanogaster. *Proc. R. Soc. B Biol. Sci.* **276**: 2625–2633.

Simmons, L.W. 2014. Sexual selection and genital evolution. *Austral Entomol.* **53**: 1–17. Wiley Online Library.

Stern, D.L., Crocker, J., Ding, Y., Frankel, N., Kappes, G., Kim, E., *et al.* 2017. Genetic and transgenic reagents for Drosophila simulans, D. mauritiana, D. yakuba, D. santomea, and D. virilis. *G3 Genes Genomes Genet.* **7**: 1339–1347.

Therneau, T., Atkinson, B. & Ripley, B. 2018. *rpart: Recursive Partitioning and Regression Trees. R Package. version 4.1-13*.

Turissini, D.A. & Matute, D.R. 2017. Fine scale mapping of genomic introgressions within the Drosophila yakuba clade. *PLoS Genet.* **13**: e1006971.

Waddington, C.H. 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**: 563.

West-Eberhard, M.J. 2003. *Developmental Plasticity and Evolution*. Oxford University Press.

Wheeler, D., Wong, A. & Ribeiro, J.M. 1993. Scaling of feeding and reproductive structures in the mosquito Aedes aegypti L.(Diptera: Culicidae). *J. Kans. Entomol. Soc.* 121–124.

Yassin, A. & David, J.R. 2016. Within-species reproductive costs affect the asymmetry of satyrization in Drosophila. *J. Evol. Biol.* **29**: 455–460. Wiley Online Library.

Yassin, A. & Orgogozo, V. 2013. Coevolution between Male and Female Genitalia in the Drosophila melanogaster Species Subgroup. *PLoS ONE* **8**: e57158.
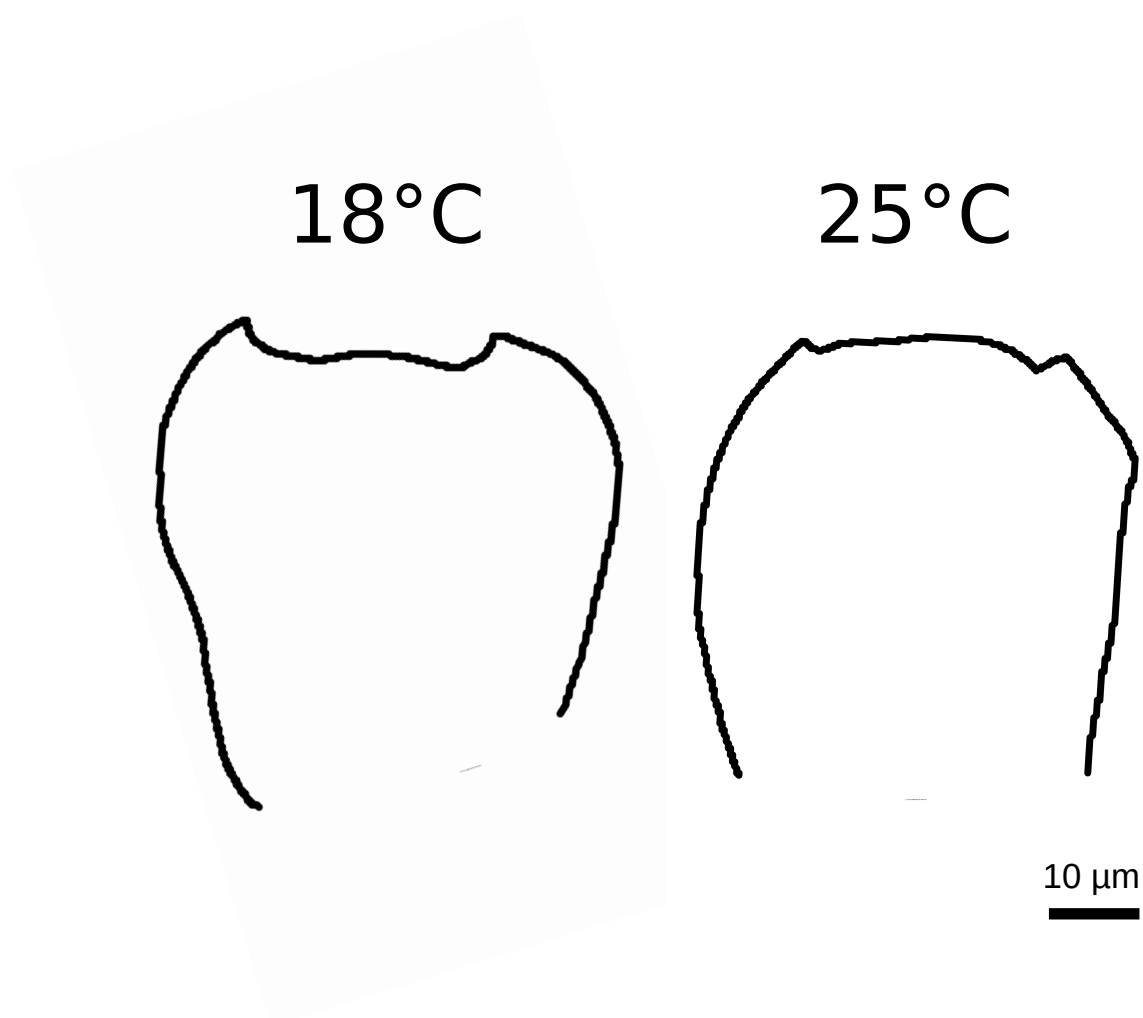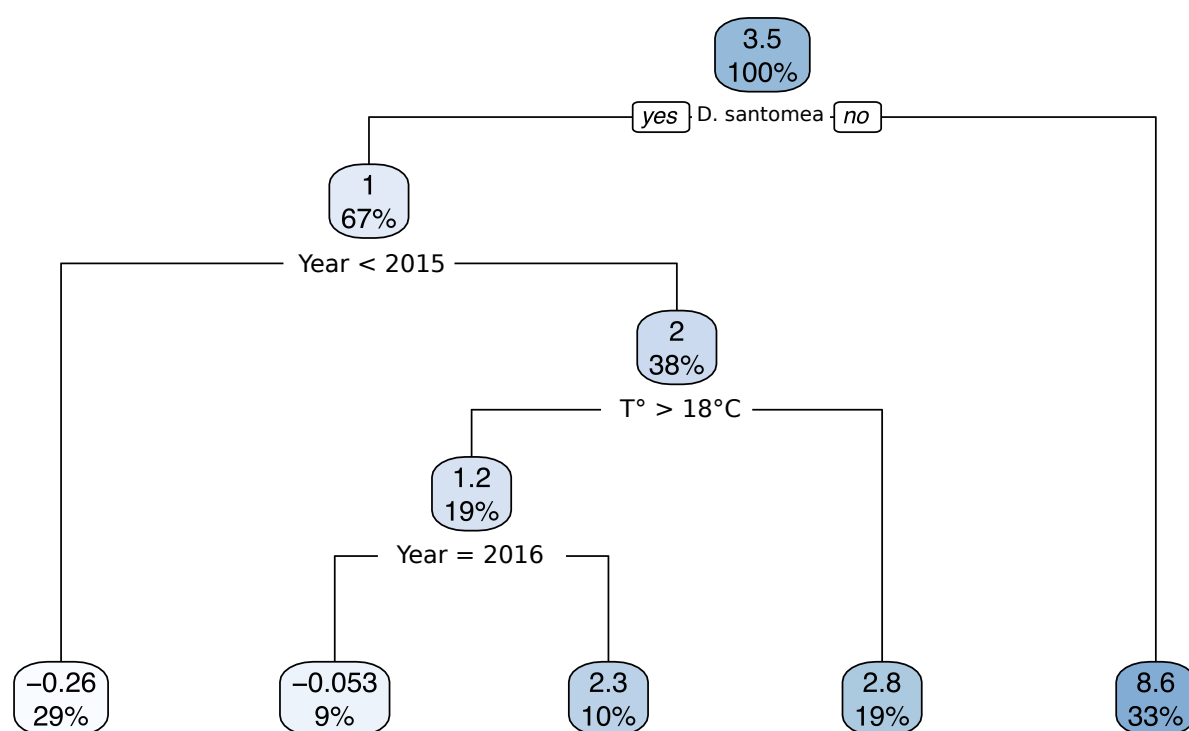
D. santomea

18°C   25°C

D. yakuba

25°C

Picture

10 μm
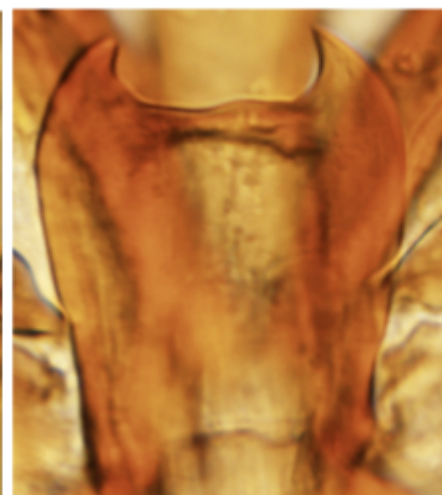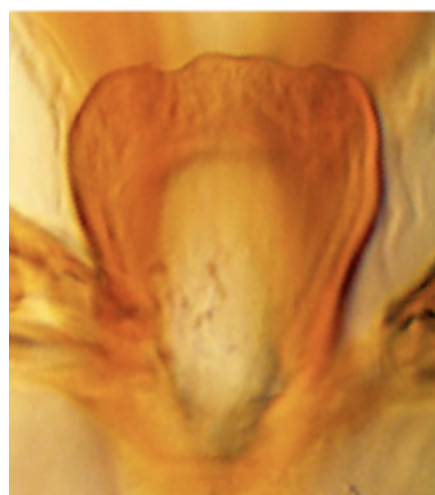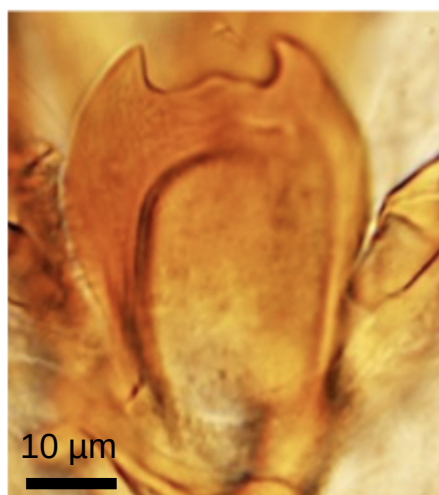
Contour
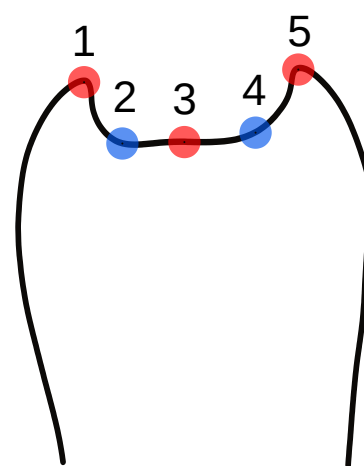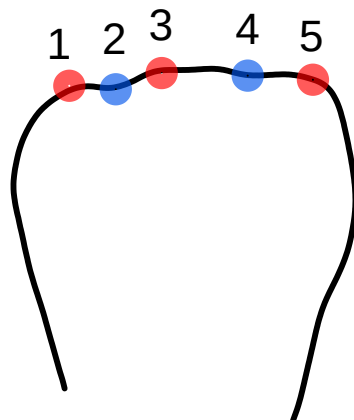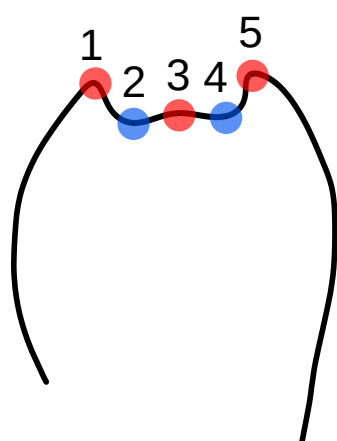
Curvature

Distance from origin

A

D. santomea

D. yakuba

1998 2001 2009 2015 2016

1970 2015 2016

B

D. santomea

D. yakuba

0.2   -1.8   -2.2   -0.5   -1.3   -2.0   -3.4

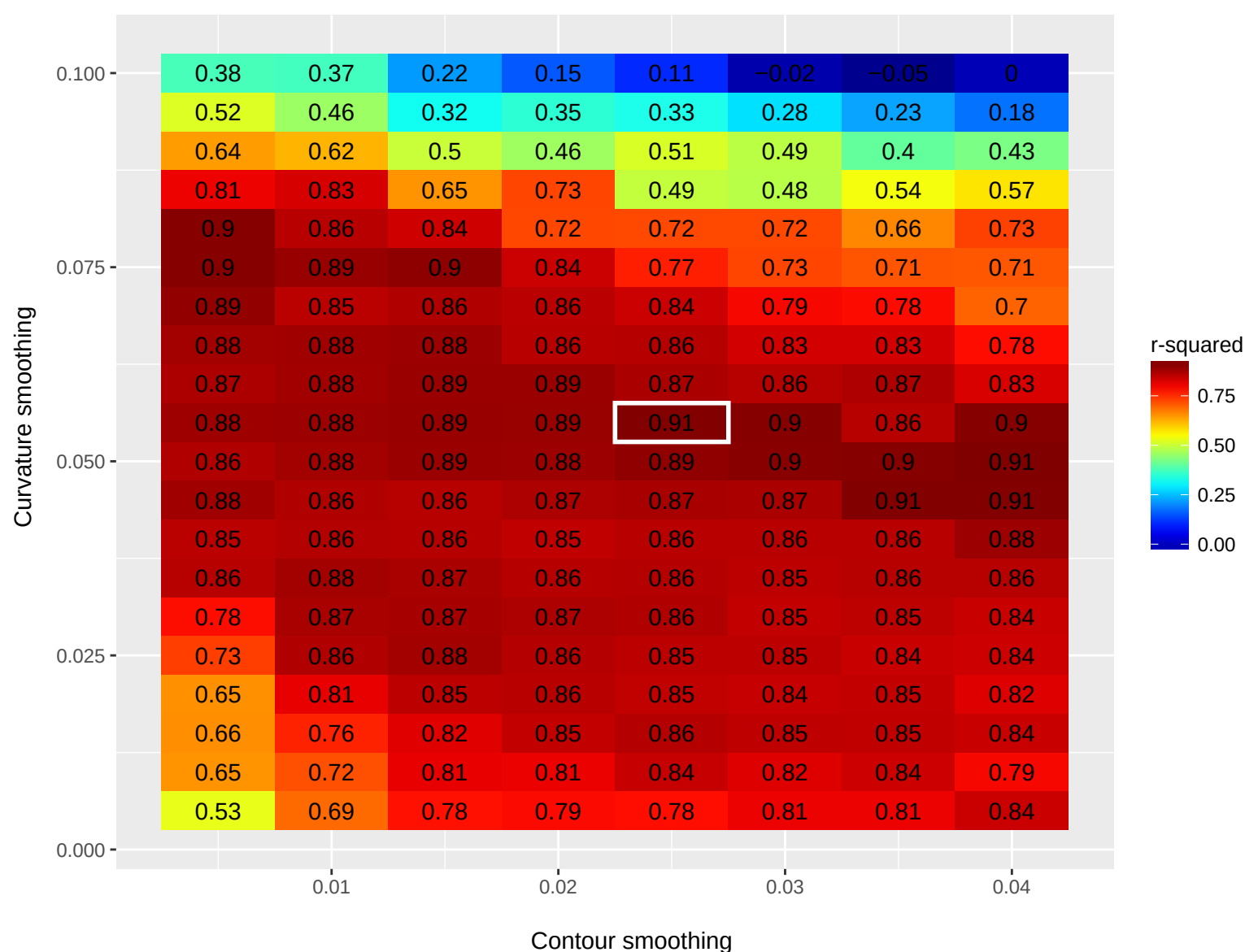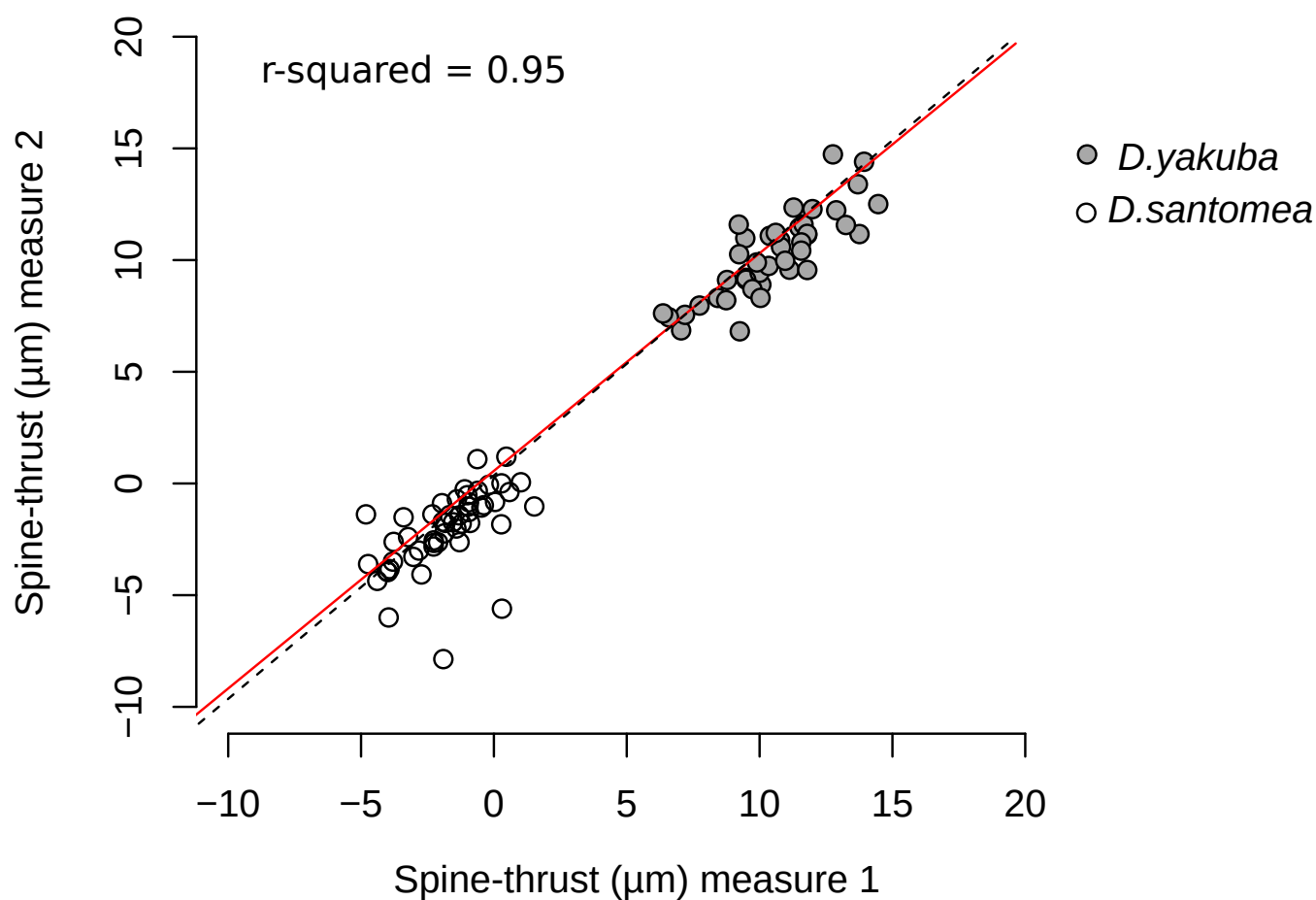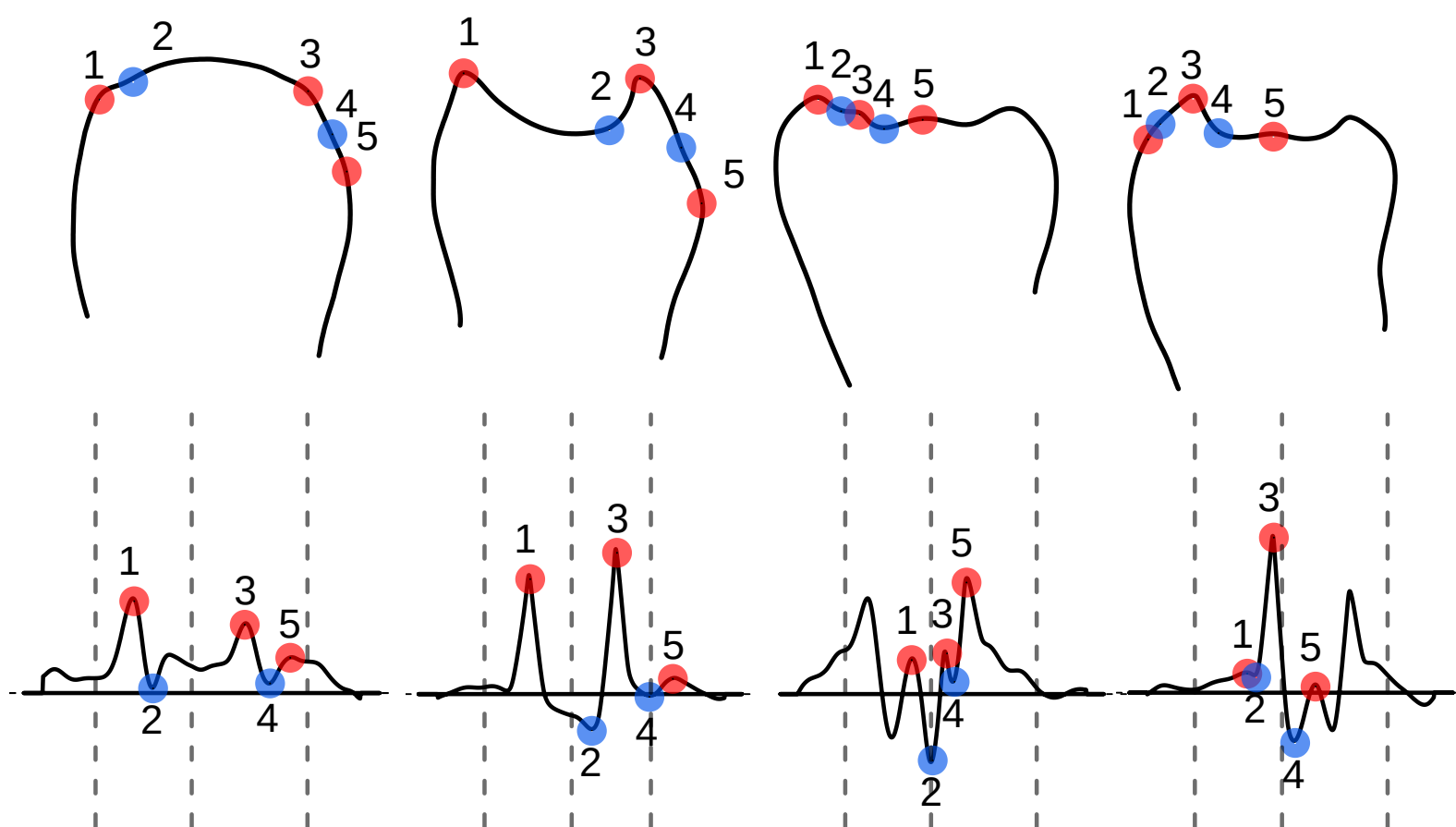0.8   0.1   -0.9   -0.1

1998 2001 2009 2015 2016

1970 2015 2016
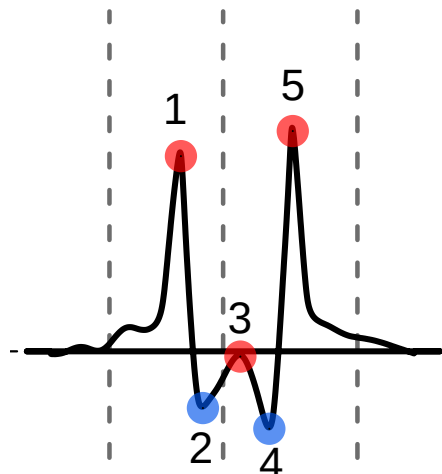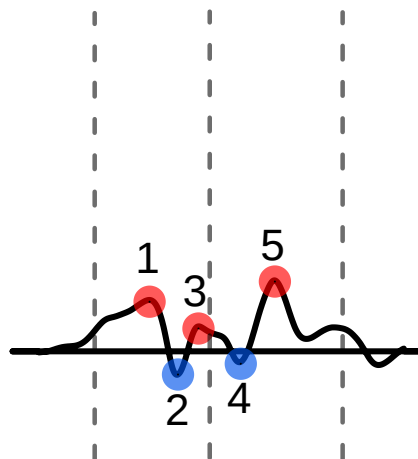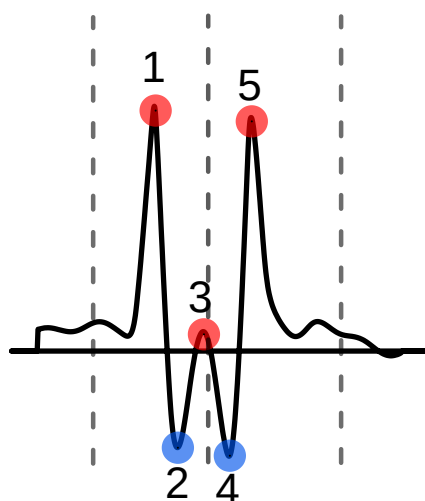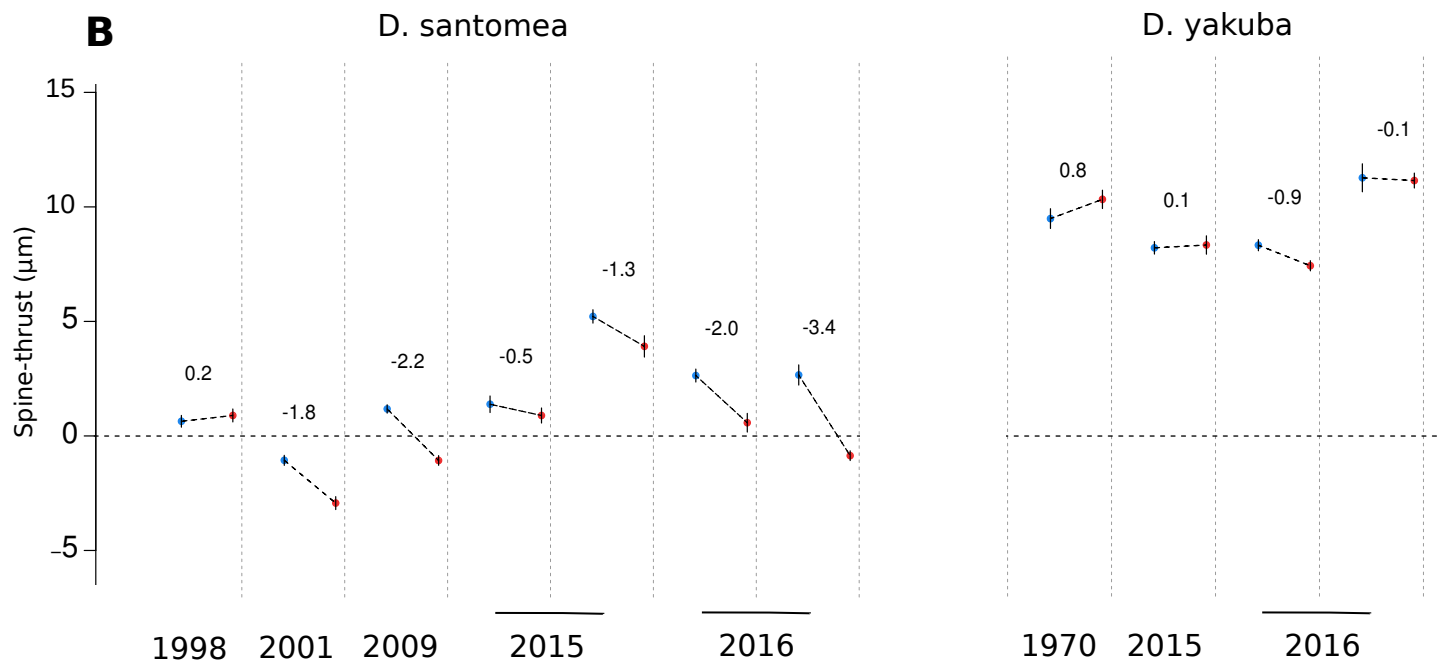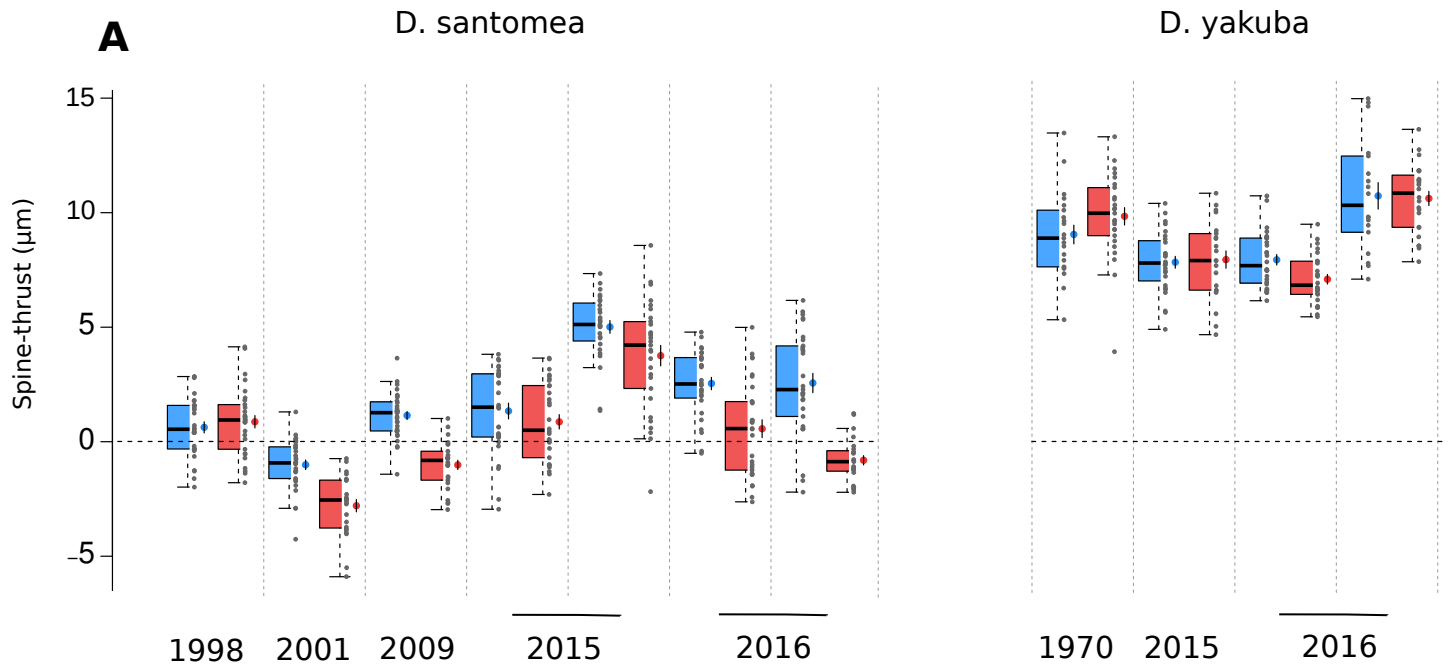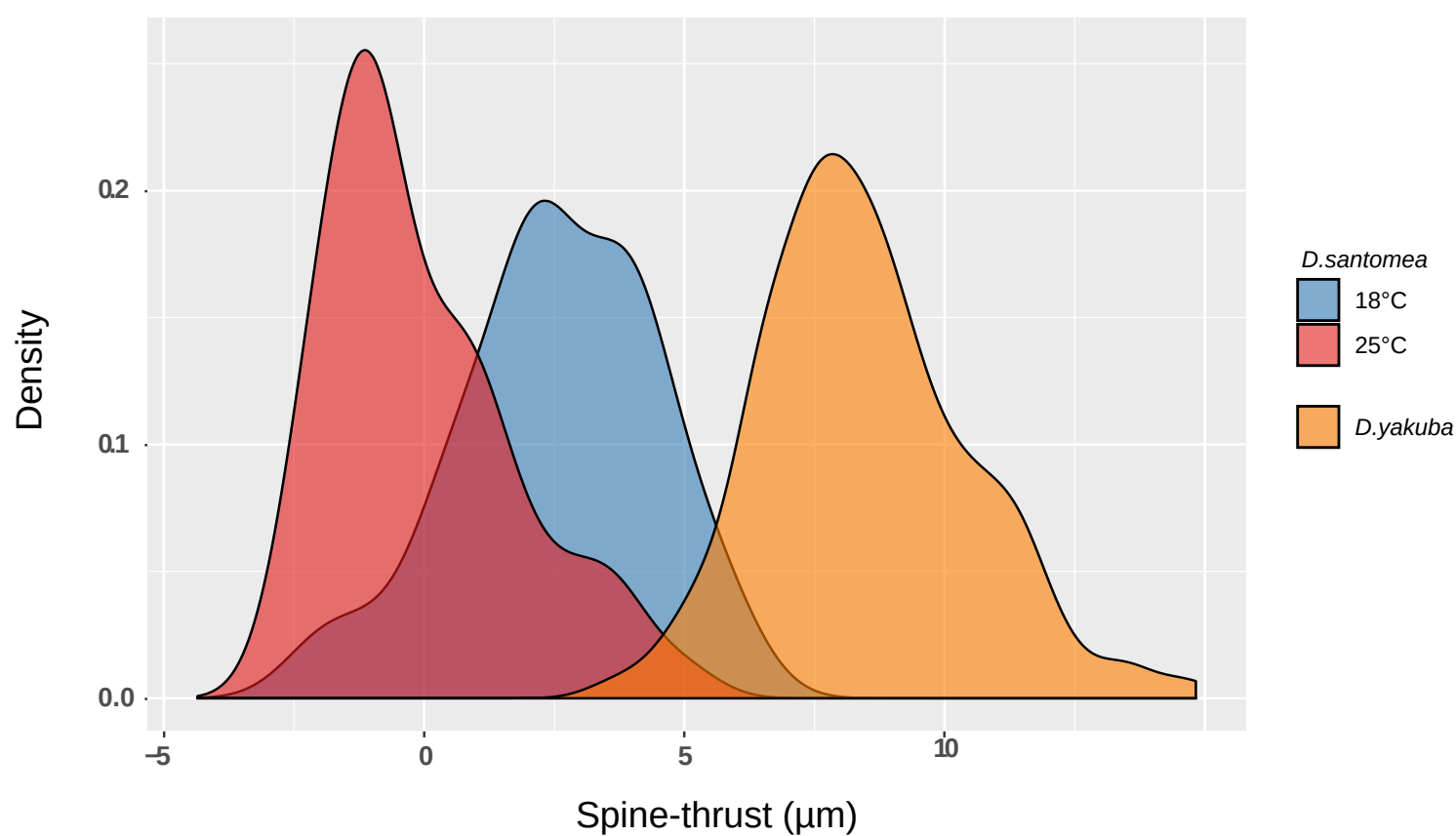
Peluffo et al. 2020

Table S1.

```
Call:
lm(formula = st ~ sp * year * temp, data = st_strains)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9035 -1.4710 -0.2689  1.3520  7.7223

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -7.505e+02  1.892e+02  -3.968 8.17e-05 ***
spYak           6.805e+02  2.138e+02   3.183  0.00154 **
year            3.760e-01  9.411e-02   3.995 7.30e-05 ***
temp            2.189e+01  8.710e+00   2.513  0.01224 *
spYak:year     -3.370e-01  1.064e-01  -3.167  0.00162 **
spYak:temp     -1.648e+01  9.816e+00  -1.679  0.09370 .
year:temp      -1.099e-02  4.333e-03  -2.536  0.01149 *
spYak:year:temp 8.301e-03  4.886e-03   1.699  0.08989 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.125 on 576 degrees of freedom
Multiple R-squared:  0.7556,    Adjusted R-squared:  0.7526
F-statistic: 254.3 on 7 and 576 DF,  p-value: < 2.2e-16
```