Manipulation of trait expression and pollination regime reveals the adaptive significance of spur length

Judith Trunschke,1,2 Nina Sletvold,1 and Jon Ågren1,3

1Plant Ecology and Evolution, Department of Ecology and Genetics, EBC, Uppsala University SE-752 36, Uppsala, Sweden
2Current Address: Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, P. R. China
3E-mail: jon.agren@ebc.uu.se

Received July 13, 2019
Accepted January 7, 2020

Understanding the mechanisms of adaptive population differentiation requires that both the functional and adaptive significance of divergent traits are characterized in contrasting environments. Here, we (a) determined the effects of floral spur length on pollen removal and receipt using plants with artificial spurs representing the species-wide variation in length, and (b) quantified pollinator-mediated selection on spur length and three traits contributing to floral display in two populations each of the short-spurred and the long-spurred ecotype of the orchid Platanthera bifolia. Both pollen receipt and removal reached a maximum at 28–29 mm long spurs in a short-spurred population visited by short-tongued moths. In contrast, pollen receipt increased linearly across the tested range (4–52 mm) and pollen removal was unrelated to spur length in a long-spurred population predominantly visited by a long-tongued moth. The experimentally documented effects on pollen transfer were not reflected in pollinator-mediated selection through female fitness or pollen removal indicating that the natural within-population variation in spur length was insufficient to result in detectable variation in pollen limitation. Our study illustrates how combining trait manipulation with analysis of causes and strength of phenotypic selection can illuminate the functional and adaptive significance of trait expression when trait variation is limited.

KEY WORDS: Flower morphology, orchid, Platanthera bifolia, pollinator-mediated selection, population differentiation, trait manipulation.

Divergent selection is a key process shaping and maintaining population differentiation and phenotypic diversity (Mitchell-Olfs et al. 2007; Siepielski et al. 2013). In flowering plants that depend on animal pollinators for successful pollen transfer, spatial variation in the composition of pollinator communities is expected to cause variation in selection on floral traits (Grant 1949; Stebbins 1970; Kay and Sargent 2009; Van der Niet et al. 2014). This is because pollinators differ in their responses to visual and olfactory signals (Huber et al. 2005; Streisfeld and Kohn 2007; Waelti et al. 2008) and in their morphological fit with the reproductive parts of the flower (Anderson and Johnson 2009; Newman et al. 2014). Comparative studies have documented correlations between floral traits and the behavior (Muchhal 2007) and morphology (Fenster et al. 2004; Anderson and Johnson 2009; Bobek et al. 2014; Wilson et al. 2017) of pollinators in several systems. However, to clarify the mechanisms responsible for such correlations, it is necessary to examine how floral traits affect pollen transfer and plant fitness in environments with contrasting pollinator assemblages (Streisfeld and Kohn 2007; Anderson and Johnson 2009; Nuismier et al. 2010; Newman et al. 2014). Such analyses will allow a test of the hypothesis that population differentiation is associated with current divergent selection on floral traits, and whether variation in selection can be attributed to differences in relationships between floral traits and pollination success in populations serviced by different pollinators.
Flower depth, that is, the depth of tube-shaped flowers or nectar spurs, influences the morphological fit between flower and pollinator and should therefore affect the efficiency of pollen removal and receipt (Darwin 1862; Nilsson 1988; Muchala and Thomson 2009). Nectar often accumulates at the bottom of flowers or floral extensions, forcing pollinators to reach deep into the flower to extract the full reward. Pollination success is expected to increase asymptotically with increasing flower depth. This is because the contact between pollinator and flower sexual parts should intensify and pollination efficiency increase only until the flower is deeper than the proboscis of the pollinator reaches (Darwin 1862; Nilsson 1988). Consequently, the minimum flower depth required for efficient pollen removal and receipt is expected to be shorter in populations visited by short-proboscis pollinators than in populations visited by long-proboscis pollinators. However, the effect of morphological fit can differ between components of pollination success: The morphological fit can be less important for pollen removal than for pollen receipt if pollen can attach to several locations on the pollinator’s body, whereas pollen deposition on a stigma requires attachment to a specific area (Cresswell 2000; Ellis and Johnson 2010).

The strength of current pollinator-mediated selection on flower depth should depend on the degree of mismatch between pollinator and floral morphology (Pauw et al. 2009; Nattero et al. 2010; Paudel et al. 2016), on available phenotypic variation in flower depth (Conner et al. 2009), and also on the magnitude of pollen limitation and opportunity for selection (variance in relative fitness; Bartkowska and Johnson 2015; Sletvold and Ägren 2016; Trunschke et al. 2017). The strength of selection on flower depth should increase with the proportion of plants in the population with a flower depth shorter than the length at which pollen transfer plateaus. Moreover, because the number of ovules produced is finite, seed production should be a saturating function of the number of pollen grains received (Waites and Ägren 2004; Richards et al. 2009). A relationship between trait expression and pollen receipt will therefore translate into pollinator-mediated selection through female function only as long as seed production is limited by pollen receipt in at least some plants. Finally, the strength and mode of net selection on flower depth is likely to be influenced also by other selective agents, and could be weaker, stabilizing, or even in the opposite direction of pollinator-mediated selection if the production of deep flowers is subject to conflicting selection (Aigner 2005; Boberg et al. 2014).

Characterizing and comparing the shape of fitness functions in different populations can be challenging for several reasons. Long-term directional or stabilizing selection is expected to reduce phenotypic variation within populations (Bulmer 1971). This means that although pollinators may have driven divergence in floral traits, current pollinator-mediated selection can be difficult to detect in natural populations (Conner et al. 2009). Moreover, trait divergence can complicate among-population comparisons of relationships between trait expression and pollination success and between trait expression and fitness. Both of these problems can be overcome by crossing divergent populations or species to obtain experimental populations that segregate for traits of interest (e.g., Schemske and Bradshaw 1999; Toräng et al. 2017), or by increasing variation through phenotypic manipulation (e.g., Andersson 1982; Sinervo et al. 1992). Experimental studies of the effects of flower depth on plant reproductive success in natural populations have typically involved the shortening of nectar spurs to produce a few categories of spur length (e.g., Nilsson 1988; Ellis and Johnson 2010). Here, we use artificial spurs to manipulate spur length in a continuous fashion allowing the assessment of the shape of the relationship between spur length and pollination success in markedly greater detail.

In this study, we examine the functional and adaptive significance of spur length in the moth-pollinated orchid Platanthera bifolia. This species displays wide variation in spur length across its range. Boberg et al. (2014) documented a positive correlation between spur length and proboscis length of the dominating local pollinator across populations in Sweden and Norway, which is consistent with pollinator-driven divergence in spur length. On the island Öland in southeastern Sweden, P. bifolia occurs as a short-spurred grassland ecotype visited by noctuid moths and hawkmoths, and a long-spurred woodland ecotype visited predominantly by the hawkmoth Sphinx ligustri, which has a proboscis that is about twice as long compared to those of most of the pollinators visiting the short-spurred ecotype (Boberg et al. 2014). Previous experiments have shown that shortening of the spur of the woodland ecotype to a length corresponding to that of the grassland ecotype reduces pollen removal, as well as pollen receipt and measures of female reproductive success in the woodland habitat (Nilsson 1988; Boberg and Ägren 2009; Trunschke et al. 2019).

Here, we used two approaches to detail the functional and adaptive significance of spur length in both grassland and woodland populations of P. bifolia on Öland. First, we experimentally determined the effect of spur length on pollen removal and receipt by replacing natural spurs with artificial spurs representing the species-wide variation in length. This allowed us not only to confirm an effect of spur length on pollination success, but also to characterize the shapes of the relationships between spur length on the one hand and pollen removal and pollen receipt on the other. Because pollinaria can attach along much of the length of the proboscis of visiting insects in P. bifolia (Nilsson 1988), whereas not all positions of attachment are likely to be conducive for transfer to stigmas of conspecific flowers (cf. Ellis and Johnson 2010), we expect pollen receipt to be more strongly related to spur length than is pollen removal. Second, we quantified current pollinator-mediated selection on spur length through pollen removal and
female reproductive success in both short-spurred grassland and long-spurred woodland populations. We tested the hypotheses that (1) the relationship between spur length and pollination success saturates at shorter spur length in the short-spurred grassland than in the long-spurred woodland populations, (2) pollen receipt is more strongly related to spur length than is pollen removal, (3) there is directional pollinator-mediated selection for longer spurs through pollen removal and female reproductive success, and the strength of selection depends on the proportion of plants that have spurs shorter than the length at which pollination success saturates, and (4) there is stabilizing net selection on spur length with shorter optimal spur length in short-spurred than in long-spurred populations.

**Material and Methods**

**PLATANTHERA BIFOLIA AND ITS ECOTYPES**

*Platanthera bifolia* (L.) Rich. is a long-lived terrestrial orchid that is widespread in the temperate regions across Eurasia (Delforge 2005). It grows in a variety of habitats ranging from grasslands, moorlands, and marshes to open woodlands, and can be found at elevations up to 2500 m above sea level. During flowering, typically between early and late June in Scandinavia, individual plants produce a single inflorescence on which flowers open sequentially from bottom to top with one to three new flowers opening per night. Flowers are white and at nighttime emit a strong scent dominated by benzenoids and linalool (Tollsten and Bergström 1989; Tollsten and Bergström 1993). The long spur produces sugar-rich nectar throughout the lifetime of the flower (Stpiczynska 1997). The species attracts a variety of sphingid and noctuid moths (Nilsson 1983; Claessens et al. 2008; Boberg et al. 2014). Pollen is packaged in sectile pollinia (two in each flower) containing hundreds of massulae (201–459 pollen grains/massulae; Nazarov and Gerlach 1997). Pollen from a given pollinium can be deposited in multiple flowers. Flowering time and morphological traits such as plant height, the number of flowers, the size of individual flowers, and the length of the nectar spur vary considerably among populations in Scandinavia (Boberg et al. 2014).

On the island Öland in southeastern Sweden, *P. bifolia* occurs as a grassland and a woodland ecotype (Boberg et al. 2014). The grassland ecotype begins to flower about two weeks later, and produces markedly shorter inflorescences with smaller flowers compared to the woodland ecotype (Boberg et al. 2014; see also Tables 1 and S1). Most noticeable is the distinct difference in spur length with flowers of the grassland ecotype having on average about 1.5 cm shorter spurs than the woodland ecotype (grassland ecotype, 21 mm [mean based on 7 population means], woodland ecotype, 36 mm [mean based on 5 population means]; Boberg et al. 2014) with very little overlap of spur lengths observed between the two ecotypes (see also Tables 1 and S1). The two ecotypes do not differ in floral scent composition (Tollsten and Bergström 1993). The difference in spur length is associated with a difference in the proboscis length of the dominating pollinators in grassland and woodland habitats: various short-proboscis moths and hawkmoths are common in grassland populations (i.e., Deilephila porcellus, *Hyles gallii*, and *Cucullia umbratica*; Nilsson 1983; Boberg et al. 2014), whereas woodland populations are predominantly pollinated by the long-proboscis hawkmoth *Sphinx ligustri* (Boberg et al. 2014).

For the present study, we selected two large woodland and two large grassland populations of *P. bifolia* in the central part of Öland. The woodland populations were located at Gråborg (coordinates in the WGS84 system, N 56.669068, E 16.598443) and Vedby (N 56.769556, E 16.645412), and the two grassland populations at Melösö (N 56.859528, E 16.856478) and Strandtorp (N 56.659951, E 16.712848). Table 1 summarizes population means (±SD) of phenotypic traits, pollen removal, and female fitness (data for the woodland population Gråborg are from Trunschke et al. 2019). Pollinator observations during the present study confirmed previously documented differences in proboscis length of pollinators in long- and short-spurred populations. In the woodland populations, *Sphinx ligustri* was the most commonly observed pollinator (proboscis length, mean ± SD, 39.1 ± 2.20 mm, N = 52; Boberg et al. 2014), whereas in the grassland populations, *Deilephila porcellus* (proboscis length of pollinators caught in the two grassland populations, 18.2 ± 0.80 mm, N = 6) and *Cucullia umbratica* (19.4 ± 0.06 mm, N = 10) were most commonly observed with some observations also of *Hyles gallii* (24.8 mm, N = 1).

**EFFECT OF SPUR LENGTH ON POLLEN REMOVAL AND RECEIPT**

To determine the shape of the relationships between spur length and pollen removal and between spur length and pollen receipt in a grassland and a woodland population, we replaced natural spurs with artificial spurs of predefined length and documented pollination success in the Gråborg (woodland) and the Strandtorp (grassland) population (Fig. 1). The length of artificial spurs ranged from 4 to 52 mm in the woodland population (13 spur length classes, with an incremental increase in length of 4 mm between classes) and from 4 to 44 mm in the grassland population (11 spur length classes). Fewer spur length classes were employed in the grassland population because of a shortage of plants, but at both sites the range of artificial spur lengths extended well beyond that of the local population (Fig. 2). Artificial spurs were made of transparent elastic surgery tubing (Tygon S-50-HL; inner diameter: 1.8 mm; outer diameter: 2.6 mm; Saint Gobain Performance Plastics, Akron, OH, USA), which was cut to the defined length and sealed at one end with fast-drying glue. The range of
Table 1. Phenotypic traits and reproductive success of open-pollinated control plants (C) and plants receiving supplemental hand pollination (HP) in two grassland and two woodland populations of the orchid *Platanthera bifolia*, on the island Öland, south-eastern Sweden in 2016. Means ± SD are given. Number of pollinia removed and female fitness (the product of number of fruits and mean fruit mass) were used to calculate relative fitness in the phenotypic selection analyses (see Methods). Pollen limitation (PL) was estimated as 1 – (mean female fitness of open-pollinated control plants/mean female fitness of plants receiving supplemental hand-pollination), and 95% confidence intervals were obtained with bootstrapping. The opportunity for selection (I) was quantified as the variance in relative pollen removal in the open-pollinated control, and as the variance in relative female fitness in the control and in the hand-pollination treatments, respectively. Pollen removal was not possible to score in a few plants, and sample sizes were therefore somewhat lower for pollen removal than for female fitness and other traits in the control in three of the four populations (Melös N = 129, Gräborg N = 118, Vedby N = 97).

| Variable                  | Grassland               |                 | Woodland               |                 |
|---------------------------|-------------------------|-----------------|------------------------|-----------------|
|                           | Melös                   | Strandtorp      | Gräborg                | Vedby           |
|                           | C (N = 141)             | HP (N = 71)     | C (N = 119)            | HP (N = 58)     |
|                           |                         |                 | C (N = 101)            | HP (N = 47)     |
| Plant height (cm)         | 17.8 ± 4.2              | 18.7 ± 3.8      | 22.8 ± 5.6             | 23.3 ± 5.0      |
|                           | 33.9 ± 6.9              | 33.2 ± 6.5      | 33.5 ± 8.6             | 31.8 ± 9.0      |
| Number of flowers         | 15.0 ± 4.6              | 14.7 ± 3.4      | 14.4 ± 3.9             | 15.0 ± 4.0      |
|                           | 18.4 ± 5.5              | 17.4 ± 6.3      | 17.7 ± 7.0             | 18.0 ± 6.7      |
| Flower size (cm²)         | 1.95 ± 0.44             | 1.92 ± 0.46     | 2.08 ± 0.51            | 2.09 ± 0.54     |
|                           | 3.94 ± 1.07             | 3.87 ± 0.89     | 4.22 ± 1.11            | 3.97 ± 1.04     |
| Spur length (mm)          | 20.5 ± 2.2              | 21.2 ± 2.3      | 20.9 ± 2.8             | 21.2 ± 2.2      |
|                           | 33.9 ± 3.7              | 33.3 ± 3.9      | 35.6 ± 3.9             | 34.7 ± 5.1      |
| Number of pollinia removed| 29.4 ± 8.9              | 28.4 ± 8.0      | 30.3 ± 11.1            | 30.5 ± 13.8     |
| Prop. of pollinia removed (%) | 98.5 ± 4.5            | 98.1 ± 5.6      | 82.1 ± 17.8            | 88.7 ± 15.0     |
| Number of fruits          | 12.4 ± 4.6              | 13.6 ± 3.7      | 11.8 ± 4.3             | 14.0 ± 4.4      |
|                           | 14.2 ± 6.3              | 16.7 ± 6.0      | 10.1 ± 7.1             | 15.3 ± 7.6      |
| Fruit set (%)             | 81.9 ± 14.7             | 92.3 ± 10.5     | 81.5 ± 19.5            | 92.8 ± 13.1     |
|                           | 75.0 ± 20.6             | 96.2 ± 5.5      | 53.7 ± 24.5            | 83.5 ± 23.5     |
| Fruit mass (mg)           | 9.46 ± 3.46             | 9.42 ± 2.96     | 9.94 ± 3.39            | 9.61 ± 3.03     |
|                           | 16.47 ± 6.98            | 17.09 ± 6.19    | 15.04 ± 6.43           | 15.08 ± 8.28    |
| Female fitness (mg)       | 123.9 ± 84.8            | 129.4 ± 58.2    | 125.4 ± 77.8           | 140.1 ± 73.2    |
|                           | 258.8 ± 205.2           | 297.1 ± 188.4   | 177.3 ± 180.2          | 251.1 ± 201.4   |
| Pollen limitation (95% CI)| 0.04 (-0.050 to 0.229)  | 0.10 (-0.117 to 0.172) | 0.13 (-0.071 to 0.298) | 0.29 (0.034 to 0.479) |
| Ipollen removal            | 0.09                    | 0.08            | 0.14                   | 0.20            |
| Ifemale fitness           | 0.47                    | 0.20            | 0.39                   | 0.27            |
|                           | 0.63                    | 0.40            | 1.03                   | 0.64            |
Figure 1. Experimental manipulation of spur length of Platanthera bifolia flowers. In three flowers per plant, spurs were replaced by artificial spurs made of silicon tubing of a given length. Artificial spur lengths covered the range of spur lengths observed in P. bifolia in Scandinavia (Boberg et al. 2014), and somewhat extended the phenotypic range observed in natural populations on Öland (12–47 mm across the four populations of the present study).

The following experimental procedure was replicated on three occasions corresponding to early, mid, and late flowering in the woodland and the grassland population. All experimental plants were covered with mosquito net shortly before their first flowers opened to prevent pollinator visits prior to the experimental manipulation. On the day before the experiment was conducted, we selected plants of similar phenological stage with at least seven open flowers (26 plants in the woodland population and 22 plants in the grassland population). On a given experimental plant, we recorded plant height with a ruler to the nearest 0.1 cm, and the vertical and horizontal maximum widths of petals, and spur length from its base to the tip with a digital caliper to the nearest 0.01 mm (second or third flower from the bottom of the inflorescence was measured), and we counted the number of flowers produced. Spur length treatment was randomly assigned to experimental plants (two plants of each spur length class on each day experimental treatments were applied, except for the last day in the grassland population when there was only one replicate per class). We calculated Pearson correlation coefficients to check that assigned artificial spur length was not correlated with any of the floral characters scored. When the experimental treatment was applied, the original spur was cut off in three fully developed flowers per plant, and replaced by an artificial spur of predefined length (same length for all experimental flowers on a given plant). The artificial spur was attached to the flower by gently sliding it over the remaining approximately 3-mm long spur basis and fixing it to the ovary with a small piece of green elastic tape (Fig. 1). We injected about 5 µL of sugar solution (sucrose) at the end tip of each artificial spur with an injection needle, resulting in a nectar column reaching about 2 mm from the bottom of the artificial spur. This nectar column height corresponds to that of a flower that has been recently depleted of nectar by a pollinator with a proboscis of almost the length of the spur, and should increase the probability that a flower visitor will reach as far into the spur as possible to reach the reward. The sugar concentration corresponded to the mean sugar concentration of nectar sampled from 15 randomly chosen plants in each of the two populations one day prior to the start of the experiment. For each plant, sugar concentration (sucrose equivalents) was determined with a refractometer (Sugar/Brix Refractometer 300003, Sper Scientific Ltd., Scottsdale, AZ, USA) in a sample pooled from three flowers (mean concentration ± SE, Gråborg: 13.5 ± 0.76% and Strandtorp: 18.1 ± 0.66%).

To document the effects of spur length on pollen removal and pollen receipt, we recorded the number of pollinia removed and the number of massulae received three days after treatment application. For each experimental flower, number of pollinia removed was scored at the end of the three-night period, whereas newly received massulae were recorded each morning. Newly received massulae can be distinguished from massulae received more than 24 h ago, because the latter are typically transparent and partly absorbed by the stigmatic surface. We calculated the total number of massulae received by each flower by summing the three records of newly received massulae.

To determine whether pollen removal and receipt in flowers with artificial spurs differed from flowers with intact spurs of similar length, we recorded pollination success also in three intact control flowers on each experimental plant following the same protocol as for manipulated flowers (only data from plants with artificial spurs within the range of natural variation were considered in this comparison). Pollination success did not differ between flowers with artificial spurs (spur length, 28–44 mm) and flowers with intact natural spurs (spur length, 26.2–44.6 mm, N = 30 plants) in the Gråborg woodland population (number of pollinia removed per flower, mean ± SD, 0.54 ± 0.54 vs. 0.59 ± 0.50, t = 0.33, P = 0.742; number of massulae received per flower, 8.7 ± 10.5 vs. 8.8 ± 7.9, t = 0.074, P = 0.941).
POLLINATOR-MEDIATED SELECTION

In 2016, we quantified selection via pollen removal and female reproductive success on spur length (expected to influence efficiency of pollen transfer) and on three morphological characters contributing to floral display and therefore expected to influence pollinator attraction (plant height, flower number, and size of...
individual flowers) in the two grassland and the two woodland populations. Results of the analysis of phenotypic selection in one of the woodland populations (Gråborg) have been published previously in Trunschke et al. (2019) and are included here to facilitate comparison with results from the other three populations. In each population, we marked up to 250 plants at the bolting stage (Gråborg, Melösa, and Strandtorp populations) or shortly after flowering had started (Vedby population; see Table 1 for number of plants included in the analysis in each population). Once at least two thirds of the flowers of a given individual had opened, we measured the four focal traits as described above.

To quantify selection mediated by pollinators, two thirds of the plants were randomly assigned to an open-pollinated control treatment (C), and one third received supplemental hand-pollination (HP). Fewer plants were included in the hand-pollination treatment because a lower variance in fitness can be expected when pollen is supplied in surplus to all plants. Plants in the hand-pollination treatment received supplemental pollination on two to three occasions during flowering, and each time, all open flowers were pollinated with pollen from a minimum of two different donors located at least 5 m away from the recipient plant.

To quantify pollen removal, we scored for each individual the total number of pollinia removed by carefully inspecting flowers with a hand lens at the end of flowering. To quantify female fitness, we counted the number of fruits produced and harvested up to three fruits per plant at fruit maturation. Fruits were brought to the lab and individually weighed to the nearest 0.01 mg, and for each individual, we estimated female fitness as the product of number of fruits and mean fruit mass.

**STATISTICAL ANALYSIS**

To characterize the functional relationship between spur length and pollination success, we first calculated for each individual the mean number of pollinia removed and the mean number of massulae received in the three flowers with artificial spurs. We fitted quadratic regressions (mean-centered polynomials) to these plant means to determine whether any curvature could be detected and to examine at which spur lengths maximum pollen removal and receipt are predicted in the two populations. To test for differences in the linear and quadratic terms between the grassland and the woodland population, we analyzed a model that in addition included site and interactions between site and the linear and quadratic terms, respectively. Furthermore, because we expect a saturating function of pollination success with increasing spur length, we also fitted a negative exponential function of the form \( y = a(1 - \exp(-bx)) \), where \( y \) is the mean number of pollinia removed per flower or the mean number of massulae received per flower, and \( x \) is the length of the artificial spur of a given plant (using the nlstools R package; Baty et al. 2015). In this model, \( a \) is an estimate of the asymptotic value of the dependent variable, and \( b \) reflects the rate at which the asymptote is reached. We used Akaike’s information criterion (AIC) to examine whether negative exponential functions fit the data better than do the quadratic functions (cf. Burnham and Anderson 2002).

In the woodland population, entire pollinia had been deposited by pollinators in eight experimental flowers. These flowers were excluded from the analysis of pollen receipt, because when whole pollinia are deposited it is unclear how many massulae actually get in contact with the stigmatic surface and contribute to pollination. To examine whether this procedure was likely to influence conclusions, we conducted a second analysis in which these flowers were included after assigning them the maximum observed value for number of massulae received. The two analyses gave qualitatively similar results, and only the former is presented below.

We used nested ANOVA to examine whether plant morphology (plant height, number of flowers, flower size, and spur length) and components of female reproductive success varied among pollination treatments, ecotypes, and populations. The ANOVA model included pollination treatment (open-pollinated control vs. supplemental hand-pollination), ecotype (grassland vs. woodland), population nested within ecotype, the pollination treatment \( \times \) ecotype, and pollination treatment \( \times \) population nested within ecotype interactions as explanatory variables.

We quantified for each population the degree of pollinator limitation of pollen removal and seed production. Pollen-removal failure was quantified as \( 1 - \text{mean proportion of pollinia removed.} \)

Pollen limitation of seed production was quantified as \( PL = 1 - (\text{mean female fitness in the open-pollinated control/mean female fitness among plants receiving supplemental hand-pollination}) \).

We calculated 95% confidence intervals (CIs) for this estimate using bootstrapping (1000 iterations; boot-package in R; Canty and Ripley 2017). For each population, the opportunity for selection through pollen removal and through female fitness in the open-pollinated control treatment was quantified as the variance in relative pollen removal (number of pollinia removed divided by mean pollen removal) and relative female fitness (female fitness divided by mean female fitness), respectively.

To quantify phenotypic selection, we used multiple regression analysis, in which relative pollen removal (one component of male reproductive success) or relative female fitness was regressed on the four standardized traits plant height, number of flowers, flower size, and spur length (Lande and Arnold 1983). Traits were standardized to a mean of zero and a standard deviation of 1, separately by population and pollination treatment, because we were interested in comparing the strength of selection among plants experiencing two different pollination regimes (open-pollinated control vs. supplemental hand-pollination; cf. De Lisle and Svensson 2017). For each population, we estimated selection through pollen removal in the open-pollinated control,
and through female fitness separately by pollination treatment. Linear selection gradients ($\beta_i$) were estimated from models that included linear terms only, whereas quadratic selection gradients ($\gamma_i$) were estimated from models that included both linear and quadratic terms. Quadratic selection gradients were quantified as twice the partial regression coefficients extracted from these models. We estimated pollinator-mediated selection through female fitness as the difference in selection gradients between the two pollination treatments ($\Delta \beta_{\mathrm{Poll}} = \beta_C - \beta_{\mathrm{HP}}, \Delta \gamma_{\mathrm{Poll}} = \gamma_C - \gamma_{\mathrm{HP}}$), where $\beta_C$ and $\gamma_C$ are selection gradients estimated for open-pollinated control plants (net selection) and $\beta_{\mathrm{HP}}$ and $\gamma_{\mathrm{HP}}$ are selection gradients estimated for plants receiving supplemental hand-pollination (nonpollinator-mediated selection; Sandring and Ågren 2009; Sletvold and Ågren 2010). To determine the statistical significance of pollinator-mediated selection through female fitness, we used ANCOVA including the four standardized traits and their interactions with pollination treatment as independent variables and relative female fitness as the dependent variable. For all models, we used the plotting and diagnostic functions of the car package within the R software (Fox and Weisberg 2011) to check for outliers, normality of residuals, and collinearity among variables. Collinearity was not problematic as all variance inflation factors were <3. Residuals were normally distributed, except for the analyses of selection through pollen removal in the two grassland populations (Melösä and Strandtorp). To verify statistical significance of estimates of selection gradients through pollen removal in these populations, we used the nonparametric bootstrapping procedure implemented in the R package boot (Canty and Ripley 2017). We estimated 95% CIs of selection gradients based on resampling with replacement (1000 iterations) and considered selection gradients statistically significant if their confidence interval did not include zero. The results were consistent with those of parametric significance tests and therefore only the latter are presented below.

All statistical analyses were performed with the statistical software R Developmental Core Team version 3.5.3 using the R Studio interface (R Core Team 2019).

**Results**

**VARIATION IN SPUR LENGTH AND OTHER TRAITS**

The two woodland populations had longer spurs compared to the two grassland populations (mean ± SD, Gråborg, 33.7 ± 3.76 mm, $N = 177$; and Vedby, 35.4 ± 4.32 mm, $N = 148$ vs. Melösä, 20.8 ± 2.26 mm, $N = 212$; and Strandtorp, 21.0 ± 2.60 mm, $N = 221$; Tables 1 and S1, Figure 2E–H). In addition, plants in the woodland populations were taller and produced larger flowers compared to plants in the grassland populations, but there was no difference in number of flowers produced (Tables 1 and S1). Traits recorded were weakly to moderately and positively correlated within populations (median [range], Pearson correlations, Gråborg 0.40 [0.21–0.54], Vedby 0.50 [0.44–0.64], Melösä 0.28 [0.07–0.51], and Strandtorp 0.30 [0.12–0.53]).

**FUNCTIONAL RELATIONSHIP BETWEEN SPUR LENGTH AND POLLINATION SUCCESS**

Quadratic regression demonstrated significant curvature to the relationship between pollination success and spur length in the grassland population. Both the linear ($b$) and the quadratic ($c$) partial regression coefficients were statistically significant in models of pollen removal ($b = 0.0117, t = 2.15, P = 0.0366; c = -0.0012, t = 2.53, P = 0.0143; R^2_{\text{adj}} = 0.14$) and receipt ($b = 0.233, t = 2.01, P = 0.0494; c = -0.0273, t = 2.63, P = 0.0111; R^2_{\text{adj}} = 0.14$). The quadratic regressions indicated that maximum pollen removal and receipt were reached at a spur length of 29 and 28 mm, respectively, which correspond to the upper end of the spur length distributions in the two grassland populations (Fig. 2E and G).

By contrast, in the woodland population pollen receipt increased with spur length ($b = 0.238, t = 3.47, P = 0.0009$), but no significant curvature was detected ($c = -0.00464, t = 0.90, P = 0.3713; R^2_{\text{adj}} = 0.12$), and there was no significant effect of spur length on pollen removal ($b = 0.0041, t = 1.02, P = 0.3094; c = -0.00050, t = 1.63, P = 0.1064; R^2_{\text{adj}} = 0.02$). The statistically nonsignificant quadratic term indicated that maximum pollen receipt would be reached at a spur length of 54 mm, that is, at a spur length beyond that of the longest artificial spurs tested and beyond the phenotypic variation observed in the woodland populations (Fig. 2F and H). The effect of spur length on pollen receipt differed significantly between the two populations (significant site × quadratic term interaction, $F_{1,127} = 4.3, P = 0.0399$), whereas the effect of spur length on number of pollinia removed did not (interactions with site, linear term, $F_{1,127} = 0.0003, P = 0.987$; quadratic term, $F_{1,127} = 1.6, P = 0.2072$; Fig. S1).

Akaike’s information criterion indicated that in both populations the negative exponential function (Fig. 2A–D) provided a slightly better fit to the effect of spur length on pollen removal than did the quadratic function (AIC, woodland, 125.5 vs. 127.9; grassland, 85.6 vs. 87.0). The same was true for the effect of spur length on pollen receipt in the woodland population (568.4 vs. 569.8), but not in the grassland population (424.8 vs. 423.6). However, the difference in fit between the negative exponential and the quadratic function was in all cases small as indicated by the small difference in AIC, and there was thus no strong support for one model over the other.

**OPPORTUNITY FOR SELECTION AND SELECTION THROUGH POLLEN REMOVAL AND FEMALE FITNESS**

The opportunity for selection through female fitness was higher than through pollen removal in all four study populations.
Figure 3. Directional selection through pollen removal (selection gradient, $\beta \pm SE$; gray bar), and net selection ($\beta_C \pm SE$; white bar), nonpollinator-mediated selection ($\beta_{HP} \pm SE$; black bar) and pollinator-mediated selection ($\Delta \beta_{Poll} \pm SE$; dashed bar) through female fitness on the four morphological characters (A) spur length, (B) number of flowers, (C) plant height, and (D) flower size in two grassland and two woodland populations of the orchid Platanthera bifolia in 2016. Statistical significance of selection gradients at $P < 0.05$ is indicated with an asterisk.

(Table 1). The proportion of pollinia removed was very high in all populations (pollen removal failure 2–18%, $N = 4$ populations; Table 1), and varied little among individuals within populations. Female fitness was significantly lower in the two grassland than in the two woodland populations, but pollen limitation of female fitness tended to be stronger in the woodland populations. Pollen limitation (95% CI) was 0.13 (–0.071 to 0.298) and 0.29 (0.034 to 0.479) in the two woodland populations and 0.04 (–0.050 to 0.229) and 0.10 (–0.117 to 0.172) in the two grassland populations (Table 1).

There was no evidence for directional or stabilizing pollinator-mediated selection on spur length through pollen removal or female fitness in any of the four study populations (Fig. 3A; Tables S2 and S3). However, there was directional selection through pollen removal for more flowers in all four populations (Fig. 3B; Table S2), and for larger flowers in the two woodland populations (Fig. 3D; Table S2). Moreover, there was directional pollinator-mediated selection through female fitness for more flowers in one grassland (Meliosa) and one woodland population (Vedby), for larger flowers in the same woodland population, and for longer spurs in one of the grassland populations (Strandtorp; Figs. 3A,C and D; Table S2).

In the hand-pollination treatment, several traits were subject to selection indicating nonpollinator-mediated selection through female fitness. There was nonpollinator-mediated selection for more flowers in all four populations (Fig. 3B; Table S2). In one grassland population (Meliosa), also the quadratic selection gradient was significant reflecting a disproportionate increase in female fitness with increasing number of flowers due to selective agents other than pollinators (Table S3). In addition, there was nonpollinator-mediated selection for taller inflorescences in one grassland (Meliosa) and one woodland population (Vedby), for larger flowers in the same woodland population, and for longer spurs in one of the grassland populations (Strandtorp; Figs. 3A,C and D; Table S2).

Net selection through female function favored production of many flowers in all populations (significant directional selection; Fig. 3B; Table S2) and female fitness increased disproportionately with increasing number of flowers in one grassland (Meliosa) and in both woodland populations (significant positive quadratic selection gradients; Table S3). In addition, there was net selection for longer spurs and taller plants in the two grassland populations, and for larger flowers in one woodland population (Vedby; Fig. 3A,C and D, Table S2).

Discussion

Understanding the mechanisms behind adaptive population differentiation requires identification of traits contributing to local adaptation and an analysis of their relationship to fitness in contrasting environments. Floral depth has been identified as a
FUNCTIONAL RELATIONSHIP BETWEEN SPUR LENGTH AND POLLINATION SUCCESS

In plants producing nectar at the bottom of tube-shaped flowers or nectar spurs, the efficiency of pollen transfer between pollinator and flower should be a saturating function of flower depth, and the depth at which saturation is reached should depend on the proboscis length of the primary pollinator (Nilsson 1988). Moreover, if pollen can attach to several locations on the pollinator’s body and pollen deposition on a stigma requires attachment to a specific area, the relationship between spur length and pollen receipt should be stronger than that between spur length and pollen removal. The results of the experiment using artificial nectar spurs partly supported these predictions. First, they indicated that pollen receipt reached a maximum at shorter spur length in the grassland population, where the mean proboscis lengths of the three most abundant pollinators are short compared to that of the main pollinator in the woodland population. In the woodland population, pollen receipt increased with increasing spur length, but no significant curvature was detected. Second, although pollen removal was predicted to reach its maximum at the same spur length as pollen receipt in the grassland population, it was not significantly related to spur length in the woodland population.

However, the variance in pollen removal and receipt was large (Fig. 1) suggesting considerable stochasticity in pollen transfer, and there was no clear evidence that pollination success was an asymptotic function of spur length: the difference in fit between the negative exponential and the quadratic models was in all cases small.

Similar to the present findings, experimental reduction in spur length tended to have a weaker effect on pollen removal than on pollen receipt in previous experiments with *P. bifolia* (Nilsson 1988; Boberg and Ågren 2009; Trunschke et al. 2019) and in the orchid *Satyrium longicauda* (Ellis and Johnson 2010), suggesting that spur length is less critical for pollen removal than it is for pollen receipt. More generally, whenever pollen receipt requires a higher degree of precision of pollen placement on the pollinator than does removal, relationships between flower morphology and pollen receipt are expected to be stronger than relationships between flower morphology and pollen removal (Cresswell 2000; Delph and Ashman 2006). However, because pollen removal is only the first step in the chain of events leading to successful pollen delivery (Minnaar et al. 2019), this should not be taken to mean that selection on spur length and other traits affecting morphological fit and efficiency of pollen transfer is necessarily stronger through female than through male function. Successful transfer of pollen to a compatible plant should depend as much on precise pollen placement on the pollinator body as does pollen receipt, and the relationship between trait expression and pollen removal need therefore not reflect the relationship between trait expression and pollen export (cf. Ellis and Johnson 2010).

POLLINATOR-MEDIATED SELECTION ON SPUR LENGTH

The relationship between spur length and pollination success documented in the artificial spur experiment was not reflected in current pollinator-mediated selection for longer spurs in any of the two grassland and two woodland populations examined. Several factors may have contributed to the lack of significant current pollinator-mediated selection on spur length in the four study populations. First, although the artificial spur experiment indicated that pollen removal (grassland population) and receipt (both populations) should increase across the range of spur length represented in the natural populations, natural within-population variation in spur length was limited compared to the range of spur lengths tested with artificial spurs making it more difficult to detect statistically significant effects on degree of pollen limitation. Second, the phenotypic selection analyses indicated that also many plants with the shortest spurs received sufficient pollen for high seed set. In the artificial spur experiment, flowers of plants with a spur length of 20 mm were predicted to receive more than four massulae on average in the woodland population (Fig. 2D). Pollinator-mediated selection through female function requires...
that a given trait influences pollen receipt, but also that the resulting differences in pollination intensity influence fitness. Third, in all four populations, rates of pollen removal were high and pollen limitation of female fitness was relatively low limiting the opportunities for selection (Sletvold and Ågren 2014; Bartkowska and Johnson 2015; Trunschke et al. 2017). However, the variance in pollen removal and pollen limitation was still sufficient for the detection of selection for larger flowers through pollen removal in the two woodland populations, and of pollinator-mediated selection through female fitness for more flowers in two populations and for taller plants in one population. Fourth, if seed production is lower after self than after cross pollination, among-plant variation in the rate of self-pollination could weaken the relationship between number of pollen grains received and female reproductive success. However, controlled crosses in *P. bifolia* did not detect any difference in fruit set or fruit size between flowers receiving self and cross pollen, respectively (Boberg and Ågren 2009), suggesting that variation in geitonogamous selfing is not likely to explain the lack of pollinator-mediated selection on spur length. Finally, a multitude of factors unrelated to pollination success, including small-scale variation in resource availability and interactions with antagonists, can affect the fate of initiated fruits and weaken the relationship between pollen receipt and female reproductive success, and thus reduce the likelihood of detecting pollinator-mediated selection. Future work should characterize the strength of the relationship between pollen receipt and seed production, explore the postpollination causes of fruiting failure in this system, and quantify spatiotemporal variation in pollinator-mediated selection and its relationship to pollen limitation.

With few exceptions, estimates of selection through female fitness were larger than the corresponding estimates through pollen removal (Fig. 3; Table S2). This can at least partly be explained by the fact that opportunity for selection was larger through female fitness than through pollen removal in all four populations (Table 1). In the two grassland populations, 98% of all pollen was removed, and in the two woodland populations as much as 82% and 89%, respectively (Table 1). Under such conditions, it is not surprising that relative number of pollinia removed was mainly a function of the number of flowers produced (Fig. 3; Table S2). In the woodland populations, number of pollinia removed was also positively related to flower size. This shows that flower size or some correlated trait not included in the study affects the likelihood of pollen removal. Flower size did not affect pollination success in a previous study where this trait was experimentally manipulated (Boberg and Ågren 2009), suggesting that the latter alternative deserves further study. Scent emission of *P. bifolia* is strong at night (Tollsten and Bergström 1989, 1993), and the distance between viscidia (the sticky attachments of the pollinaria) affects how pollinaria attach to the pollinator (cf. Maad and Nilsson 2004). It would be interesting to determine whether these traits are subject to current pollinator-mediated selection (cf. Chapurlat et al. 2019), and whether a correlation between these traits and flower size could explain the apparent selection for larger flowers through pollen removal.

We did not detect any stabilizing net selection on spur length. Instead, there was significant net directional selection for longer spurs in the two grassland populations. This suggests that benefits associated with the production of long spurs were not balanced by any detectable costs affecting female fitness, and that mean spur length in the grassland populations is lower than the optimal value in that environment. Moreover, selection for longer spurs was largely driven by selective agents other than pollinators (Fig. 3). Selection on correlated characters not included in the analysis could cause significant net selection on spur length. For example, if spur length is positively correlated with ovule number, longer spurs would be associated with higher female fitness if seed production is not pollen limited (Alexandersson and Johnson 2002). Consistent selection for longer spurs in grassland populations should over time reduce the difference in spur length between the two spur length ecotypes. However, response to the documented selection can be slowed down or prevented by several factors including temporal variation in the direction of selection (Schemske and Horvitz 1989; Young 2008; Siepielski et al. 2009), costs associated with long spurs expressed through other components of fitness than female fitness in the year of flowering (Boberg et al. 2014), and genetic constraints in the form of limited genetic variation within populations, or genetic correlations among traits influencing fitness.

Taken together, the results illustrate how a combination of trait manipulation and analysis of strength and causes of selection can throw light on both the functional and adaptive significance of trait variation within and among populations. Experimental manipulation of trait expression is a powerful means to examine the adaptive significance of a given trait and has been used in a wide range of systems (e.g., Andersson 1982; Nilsson 1988; Sinervo et al. 1992). Typically, a few different size classes are included in such experiments. Here, we produced a more continuous variation in phenotype covering the species-wide range in trait values, which allowed us to characterize the shapes of the relationships between trait expression and measures of pollination success in two contrasting environments. By comparing the shapes of such functional relationships to available phenotypic variation within a population, it is possible to predict the mode and strength of selection exerted by different agents.

**AUTHOR CONTRIBUTIONS**

JT, NS, and JA planned and designed the study. JT performed the fieldwork, analyzed the data, and wrote the first draft of the manuscript. All authors contributed to revisions.
ACKNOWLEDGMENTS
We thank L. Dück for her help in the field, and M. Larsson, T. Johansson, B. Johansson, B. Karlsson, and G. Johansson for allowance to work on their land. This study was financially supported by grants from the Swedish Research Council (JÅ), Swedish Research Council Formas (NS), and the Swedish Phytogeographical Society (JT).

DATA ARCHIVING
The data used in statistical analyses presented in this paper are available at Dryad https://doi.org/10.5061/dryad.f1vhmgx.

LITERATURE CITED
Aigner, P. A. 2005. Variation in pollination performance gradients in a *Dudleya* species complex: can generalization promote floral divergence? Funct. Ecol. 19:681–689.
Alexandersson, R., and S. D. Johnson. 2002. Pollinator-mediated selection on flower tube length in a hawkmoth-pollinated *Gladiolus* (Iridaceae). Proc. R. Soc. Lond. B 269:631–636.
Anderson, B., and S. D. Johnson. 2008. The geographical mosaic of coevolution in a plant-pollinator mutualism. Evolution 62:220–225.
Anderson, M. 1982. Female choice selects for extreme tail length in a widowbird. Nature 299:818–820.
Barlow, E., and J. Ågren. 2009. Despite their apparent integration, spur length but not perianth size affects reproductive success in the moth-pollinated orchid *Platanthera bifolia*. Funct. Ecol. 23:1022–1028.
Boberg, E., R. Alexandersson, M. Jonsson, J. Maad, J. Ågren, and L. A. Nilsson. 2014. Pollinator shifts and the evolution of spur length in the moth-pollinated orchid *Platanthera bifolia*. Ann. Bot. 113:267–275.
Bulmer, M. G. 1971. The effect of selection on genetic variability. Am. Nat. 105:201–211.
Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. Springer-Verlag, New York.
Canty, A., and B. D. Ripley. 2017. Boot: bootstrap R (S-Plus) functions. R package version 1.3-20.
Chapurlat, E., J. Ågren, J. Anderson, M. Friberg, and N. Sletvold. 2019. Conflicting selection on floral scent emission in the orchid *Gymnadenia conopsea*. New Phytol. 222:2009–2022.
Claessens, J., B. Gravendeel, and J. Kleynen. 2008. *Cucullia umbratica* L. als Bestäuber von *Platanthera x hybrida* Brugg. in Süd-Limburg (Niederlande). J. Eur. Orch. 40:73–84.
Conner, J. K., H. F. Sahl, and K. Karoly. 2009. Tests of adaptation: functional studies of pollen removal and estimates of natural selection on anther position in wild radish. Ann. Bot. 103:1547–1556.
Cresswell, J. E. 2000. Manipulation of female architecture in flowers reveals a narrow optimum for pollen position. Ecology 81:3244–3249.
Darwin, C. M. A. 1862. The various contrivances by which orchids are fertilized by insects. University of Chicago Press, Chicago.
Delforge, P. 2005. Orchids of Europe, North Africa and the Middle East. A&C Black Publishers Ltd., Lond.
ginger (Roscoea purpurea) and a tabanid fly in the Nepalese Himalayas. New Phytol. 211:1402–1411.

Pauw, A., J. Stoffberg, and R. J. Waterman. 2009. Flies and flowers in Darwin’s race. Evolution 63:268–279.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available via https://www.R-project.org/.

Richards, S. A., N. M. Williams, and L. D. Harder. 2009. Variation in pollination: causes and consequences for plant reproduction. Am. Nat. 174:382–398.

Sandring, S., and J. Ågren. 2009. Pollinator-mediated selection on floral display and flowering time in the perennial herb Arabidopsis lyrata. Evolution 63:1292–1300.

Schemske, D. W., and C. C. Horvitz. 1989. Temporal variation in selection on a floral character. Evolution 43:461–465.

Schemske, D. W., and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (Mimulus). Proc. Natl. Acad. Sci. USA 96:11910–11915.

Siepielski, A. M., J. D. DiBattista, and S. M. Carlson. 2009. It’s about time: the temporal dynamics of phenotypic selection in the wild. Ecol. Lett. 12:1261–1276.

Siepielski, A. M., K. M. Gotanda, M. B. Morrissey, S. E. Diamond, J. D. DiBattista, and S. M. Carlson. 2013. The spatial patterns of directional phenotypic selection. Ecol. Lett. 16:1382–1392.

Sinervo, B., P. Doughty, R. B. Huey, and K. Zamudio. 1992. Allometric engineering: a causal analysis of natural selection on offspring size. Science 258:1927–1930.

Sletvold, N., and J. Ågren. 2010. Pollinator-mediated selection on floral display and spur length in the orchid Gymnadenia conopsea. Int. J. Pl. Sci. 171:999–1009.

———. 2014. There is more to pollinator-mediated selection than pollen limitation. Evolution 68:1907–1918.

———. 2016. Experimental reduction in interaction intensity strongly affects biotic selection. Ecology 97:3091–3098.

Stebbins, G. L. 1970. Adaptive radiation of reproductive characteristics in angiosperms, 1: Pollination mechanisms. Annu. Rev. Ecol. Syst. 1:307–326.

Stpiczynska, M. 1997. The structure of nectary of Platanthera bifolia L. Orchidaceae. Acta Soc. Bot. Pol. 66:5–11.

Streifeld, M. A., and J. R. Kohn. 2007. Environment and pollinator-mediated selection on parapatric floral races of Mimulus aurantiacus. J. Evol. Biol. 20:122–132.

Tollsten, L., and J. Bergström. 1989. Variation and post-pollination changes in floral odours released by Platanthera bifolia (Orchidaceae). Nord. J. Bot. 9:359–362.

Tollsten, L., and L. G. Bergström. 1993. Fragrance chemotypes of Platanthera (Orchidaceae)—the result of adaptation to pollinating moths? Nord. J. Bot. 13:607–613.

Toräng, P., L. Vikström, J. Wunder, S. Wötzel, G. Coupland, and J. Ågren. 2017. Evolution of the selfing syndrome: anther orientation and herkogamy together determine reproductive assurance in a self-compatible plant. Evolution 71:2206–2218.

Trunschke, J., N. Sletvold, and J. Ågren. 2017. Interaction intensity and pollinator-mediated selection. New Phytol. 214:1381–1389.

———. 2019. The independent and combined effects of floral traits distinguishing two pollination ecotypes of a moth-pollinated orchid. Ecol. Evol. 9:1191–1201.

Van der Niet, T., R. Peakall, and S. D. Johnson. 2014. Pollinator-driven ecological speciation in plants: new evidence and future perspectives. Ann. Bot. 113:199–211.

Waelti, M. O., J. K. Muhlemann, A. Widmer, and F. P. Schiestl. 2008. Floral odour and reproductive isolation in two species of Silene. J. Evol. Biol. 21:111–121.

Wailes, A. R., and J. Ågren. 2004. Pollinator visitation, stigmatic pollen loads and among-population variation in seed set in Lythrum salicaria. Ecology 92:512–526.

Whittall, J. B., and S. A. Hodges. 2007. Pollinator shifts drive increasingly long nectar spur in column flower. Nature 447:706–712.

Wilson, T. C., B. J. Conn, and M. J. Henwood. 2017. Great expectations: Correlations between pollinator assemblages and floral characteristics in Lamiaceae. Int. J. Plant Sci. 178:170–187.

Young, H. 2008. Selection on spur shape in Impatiens capensis. Oecologia 156:535–543.

Supporting Information
Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Empirically fitted quadratic regressions of (a) number of pollinia removed per flower, and (b) number of massulae received per flower on artificial spur length in the grassland and the woodland population, respectively.

Table S1. Analysis of variance of the effects of ecotype, population nested within ecotype, and pollination treatment on phenotypic traits (plant height, number of flowers, flower size, and spur length) and components of female reproductive success (number of fruits, fruit set [proportion of flowers forming a mature fruit], fruit mass, and female fitness [total fruit mass per plant]), and of ecotype and population nested within ecotype on pollen removal.

Table S2. Directional selection gradients (±SE) for four phenotypic traits estimated based on variation in relative pollen removal and relative female fitness, respectively.

Table S3. Quadratic selection gradients (±SE) for four phenotypic traits estimated based on variation in relative pollen removal and relative female fitness, respectively.