Biological Journal of the Linnean Society, 2005, **84**, 755–765. With 7 figures

Genetics of host-plant preference in the comma butterfly *Polygonia c-album* (Nymphalidae), and evolutionary implications

SÖREN NYLIN[1]*, GEORG H. NYGREN[1], JACK J. WINDIG[2], NIKLAS JANZ[1] and ANDERS BERGSTRÖM[1]

[1]*Department of Zoology, Stockholm University, SE-106 91 Stockholm, Sweden*
[2]*Institute for Animal Science and Health (ID-Lelystad), Department Genetics and Reproduction, PO Box 65, 8200 AB Lelystad, The Netherlands*

Received 16 December 2003; accepted for publication 20 June 2004

In the Lepidoptera, sex-linked genes have been found to be of importance for species differences in, for example, host-plant preference, and have been implicated in ecological speciation. Variation within species is typically not sex-linked. However, in the comma butterfly *Polygonia c-album* (Nymphalidae) an X-linked gene has been found to play a major role in determining differences in host-plant use between two well separated populations. For this reason, we studied the role of sex-linked genes for host-plant preference within a single Swedish population of this species. Three generations of females with known pedigrees were studied in the laboratory, and they were given a choice between *Urtica dioica* and *Salix caprea* in flight cages. We found strong variation among females and significant genetic variance for host-plant preference, but no evidence for major importance of sex linkage of host-plant preference on this local scale. To what extent the observed genetic variation was due to additive genes and/or effects of major genes was not clear from the maximum likelihood analysis. In a follow-up study we sampled females over a larger area. We found strong variation among females, but not among localities, suggesting an open population structure with strong gene flow. From the combined stock, a selection experiment was performed over 2 years and six generations. The selection lines diverged after the first generation of selection and remained separate, but did not diverge further, suggesting a low degree of narrow-sense heritability and that the genetic differences may be effects of major genes. We discuss these results in relation to the possible role of genetics in the radiation of the Lepidoptera and other phytophagous insects. © 2005 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2005, **84**, 755–765.

ADDITIONAL KEYWORDS: insect – learning – REML – sex-linkage – speciation.

INTRODUCTION

Phytophagous insects represent a large proportion of the Earth's biodiversity. It has been shown that phytophagous insect taxa are consistently more species-rich than are their sister taxa (Mitter, Farrel & Wiegmann, 1988), indicating that the evolution of phytophagy has promoted adaptive radiation. One prominent example is the Lepidoptera, the second-largest insect order and thus an important model case for the role of phytophagy in ecological speciation.

An increased understanding of the genetic basis of differences in host-plant preference within and between species is one of the pieces needed to complete the puzzle. Despite the popularity of butterflies as models of evolution of insect–plant interactions ever since the seminal paper by Ehrlich & Raven (1964), there is limited evidence on the genetics of butterfly host-plant choice (Thompson & Pellmyr, 1991; but see, for example, Ng, 1988; Thompson, 1988). Indeed, the information is surprisingly scarce from phytophagous insects in general (Bernays & Chapman, 1994), especially considering the economic importance of many such insects. Exceptions include references in Courtney, Chen & Gardner (1989) and Via (1990), as well as the extensive work on aphids (for example, Hawthorne & Via, 2001).

---

*Corresponding author. E-mail: soren.nylin@zoologi.su.se

More than a decade ago Jaenike (1989) pointed out the interesting discrepancy between genetic determination of species differences in butterflies as compared with within-species differences suggested by the limited data available (see also later reviews by Thompson & Pellmyr, 1991 and Bernays & Chapman, 1994). Thompson (1988) had shown that preference differences between closely related species of *Papilio* butterflies are decided to a large degree by a gene on the X chromosome. In butterflies, females are the heterogametic sex, and hence the X chromosome is always inherited from the male parent. Female oviposition behaviour in population crosses was found to reflect the male source population when *Papilio* species were reciprocally crossed (see also Scriber, 1994 and Scriber, Giebink & Snider, 1991 for another example from *Papilio*). Sperling (1994) and Prowell (1998) noted that this seems to be a more general pattern: ecological, behavioural and physiological species differences in the Lepidoptera seem to be controlled disproportionately often by X-linked genes. The reason or reasons for this bias are not clear, but may involve one or several of the following: faster evolution of sex-linked traits in general (Charlesworth, Coyne & Barton, 1987), a special role of sex-limited traits (such as host preference) in speciation, a special role of female heterogamety, and/or physical linkage on the X chromosome facilitating the formation of adaptive gene complexes (Jaenike, 1989; Scriber, 1994; Sperling, 1994; Prowell, 1998; Ellers & Boggs, 2002). In contrast, most within-species studies available do not suggest any sex-linked genetic control of variation in host-plant use, but this has been investigated explicitly only rarely. There are, however, some examples of sex-linked genetic differences between 'host strains', 'pheromone strains' or 'races' within species of Lepidoptera for other traits that may be important in facilitating speciation for example, voltinism or pheromone response (Prowell, 1998).

Jaenike (1989) ended his review by stating that it would be very interesting to carry out reciprocal crosses among differentiated populations within species, to examine the influence of the X chromosome in bringing about intraspecific variation. Such studies could provide 'a clue as to whether the genetic determination of variation in host specificity within species are of the same sort as those that bring about differences between species' (Jaenike, 1989). A study of this sort was eventually performed by Janz (1998) for an important model species: the polyphagous comma butterfly *Polygonia c-album* L. (Nymphalidae). This species has been the focus of a series of studies investigating many different aspects of host-plant preference and offspring performance (Nylin, 1988; Nylin & Janz, 1993; Janz, Nylin & Wedell, 1994; Nylin & Janz, 1996; Janz & Nylin, 1997; Wedell, Nylin & Janz, 1997; Janz, 1998). The Swedish population of *P. c-album* is known to feed as larvae on herbs, bushes and trees in seven plant families, most of which are not closely related (Nylin, 1988), whereas the English population is much more specialized on *Urtica dioica* (Nylin, 1988; Janz & Nylin, 1997; Janz, 1998).

The gene or genes responsible for causing the difference in degree of specialization on *U. dioica* (when females were given a choice between this species and *S. caprea*) between females originating from Sweden and those form England was indeed found to reside on the X chromosome (Janz, 1998). This was the first indication of a possible evolutionary link between inter- and intraspecific variation in host-plant utilization, involving sex chromosomes. However, interspecific hybrid studies are still lacking from *Polygonia*, as are studies on genetic variation in host preference within local populations. Interestingly, larger variation among hybrid females was seen in crosses in which the male originated from Sweden compared with crosses in which the male originated from England, whereas the source of females did not seem to influence the variation (Janz, 1998). This could suggest a role for the X chromosome in within-population genetic variation, completing the evolutionary sequence from genetic variation among individuals on a local scale to species differences (Janz, 2003). In this paper we report on two studies with the aim of investigating this possibility, along with a general investigation of genetic control of host-plant preference in *P. c-album*.

Earlier studies have shown the existence of a hierarchy of host-plant preference in *P. c-album* where, on average, hosts in the order Urticales (Urticaceae: *U. dioica*; Ulmaceae: *Ulmus glabra;* Cannabidaceae: *Humulus lupulus*) are the most preferred by ovipositing females (Nylin, 1988; Janz *et al*., 1994). This is true for the Swedish population as well as for populations from England (Nylin, 1988) and Estonia (S. Nylin & N. Janz, unpubl. data), suggesting an evolutionarily conservative pattern similar to the one seen in *P. zeliacon* (Wehling & Thompson, 1997). The hosts in Urticales are also the best hosts in terms of most aspects of offspring performance, such as permitting a high growth rate and short larval development time. At the same time, however, considerable phenotypic variation in preference exists among females of the Swedish population; some females are more prone to accepting low-ranked hosts such as *Betula pubescens* than are others, and some actually seem to prefer *Salix caprea* over the hosts in Urticales (Janz *et al*., 1994).

It has been suggested that the variation in Sweden can be explained by a combination of two observations. Firstly, the fitness premium on a high larval growth rate is not as high in Sweden as it is, for instance, in England, where females are more specialized on Urticales. This is because only a single generation per year is possible in the field in Sweden and there is ample time for the successful development of this single generation up to the hibernating adult stage. This creates a situation where selection favouring specialization on 'fast' host plants is relaxed (Nylin, 1988; Scriber & Lederhouse, 1992; see also Janz *et al.*, 1994). Secondly, there are aspects of total offspring fitness other than larval development rate that actually favour the use of *S. caprea* over the use of, for example, *U. dioica*, namely larger adult size and higher fecundity of females reared on the former plant (Janz *et al.*, 1994; Nylin, Janz & Wedell, 1996). It seems plausible that genetic variation in host-plant preference could be maintained in such a situation, where trade-offs exist between different aspects of fitness (cf. Krainacker, Carey & Vargas, 1987). In addition, *P. c-album* is a strong flyer (but not migratory), so gene flow should be high and contribute to genetic variation at any given site. However, earlier phenotypic studies could not rule out, for example, the influence of learning rather than genetics causing the observed variation among females (Janz *et al.*, 1994). A secondary aim of our study was to clarify the relative roles of genetics and the experience of females in the field.

MATERIAL AND METHODS

STUDY 1

Stock used, preference studies and rearing

Females that had mated in the wild were caught in early May 1998 in the area between Åkersberga and Riala, north of Stockholm, over an approximately 20-km distance. It was necessary to collect over this relatively large area in order to obtain a large enough number of parental females; even so we obtained only nine females, one of which laid eggs that did not hatch. For the purposes of this study we regarded all females as belonging to a single population, since *P. c-album* is a strong flier and mixing of individuals from different breeding localities is likely to occur, especially in autumn before adult hibernation. This assumption was later supported by the results of Study 2. *P. c-album* is a potentially polyandrous species, but collecting was done at the very beginning of the flight season. Nevertheless, some females may have mated more than once, in which case we may have underestimated the genetic variation.

During preference trials, the females were kept singly in cages (0.5 × 0.5 × 0.5 m) with a transparent top and front. Light came from a single light bulb over the centre of the cage, using a light : dark photoperiod of 8 : 16 h. Food (a solution of sugar in water) was provided on a sponge in the middle of the cage. The bottom of the cage was covered with moist tissue paper. The host plants (roughly equally sized *U. dioica* and *S. caprea*) were presented in bottles of water on each side of the central food and light sources. Their relative position in the cages was random (determined by dice); all plants were rotated among all cages but changed whenever they showed signs of wilting. Eggs on each of the plant species were counted daily, but days when a female laid very few (fewer than ten) eggs were excluded. Each female was followed for at least 5 days if it did not die earlier (in one case only 2 days could be included). For the residual error maximum likelihood (REML) analyses the average proportions of eggs laid on *U. dioica* by each female over the whole period (calculated from the daily proportions) were used as data points. For analysis of within-generation variation (using ANOVA or a non-parametric equivalent) each day of oviposition was used as a data point in order to determine whether females differed consistently in their oviposition patterns.

Twenty larvae from each of eight parental females were reared to form the F_1 generation, following a procedure known to produce the lightly coloured summer form, which does not enter adult hibernation diapause (Nylin, 1989, 1992). This procedure consisted of a change from low temperature and short day (18 °C, 12 h light) to high temperature and long day (22 °C, 21 h light) when larvae are at the beginning of the fourth, penultimate, instar. All larvae in this study were reared on *U. dioica.* Pupal weight and larval development time was noted, and pupae were sexed according to the appearance of the future genital openings. Adults were marked individually and placed in larger cages to be mated, with access to external light from windows as well as supplementary artificial light with a UV component. Offspring of two females were placed in each cage, to obtain a mixed full sib/half sib design in the next generation. There were several males and females of each family in each cage, and we collected matings until we had achieved the desired design, returning mated males to the cages until ten males had mated twice. This design also allowed us to determine whether matings between full sibs were avoided, and if offspring of such matings are viable.

Matings occured in the afternoon and almost invariably lasted well into the night, so mating pairs were collected each evening. To make this possible and still present the butterflies with a long day (to ensure that adults did not enter reproductive diapause) a 12-h light photoperiod was combined with a 10-min period of light in the middle of the night. In many insects the length of the dark phase is used to determine the photoperiod (Beck, 1980), and this procedure was used in order to partition the dark phase into two 6-h segments. Mated females (35 in total) were used for preference studies as in the parental generation and males were returned to the cages until they had mated twice in order to produce half sibs.

For the F_2 generation, the aim was to rear offspring of ten males that had mated twice, i.e. offspring of 20 females, as in the F_1 generation. However, one mated female did not lay viable eggs, and as a result only 19 families were reared. Fifteen larvae from each female were reared. Since the offspring of this generation were not to be used further no specific breeding design was necessary, so adults were mated in arbitrarily assigned cages just to produce ovipositing females (otherwise the procedure was as above). Mated females were used in preference studies as before, with the exception that *U. dioica* was presented about 1 dm lower in the cage (further from the light source) than was *S. caprea*. The proportion of eggs laid on *U. dioica* is known to increase strongly according to the progress of the season (Nylin, 1988; Janz, 1998). This had become evident from the difference between the parental and F_1 generations, and we expected that it would not be possible to score variation in preference among females without this change in the presentation of plants, as most females would then lay close to 100% of their eggs on *U. dioca.*

Statistics

Statistical analyses, with the exception of the formal genetic analysis (next section) were performed using Statistica 5.1. Preference values were not normally distributed, as they are constrained to be between 0 and 100%. ArcSin-transformation improved normality and homogeneity of variances, but not completely. When possible (number of categories not exceeding ten), we give results both from parametric ANOVAs (general linear model, GLM) and non-parametric Kruskal–Wallis ANOVAs.

Genetic analysis

The dataset consisted of three generations: (grand)mothers in Generation 1 (P), full sib families/parents in Generation 2 (F_1), and their offspring, a mixture of full and half sib families in Generation 3 (F_2). Conventional offspring–parent regressions or half sib/full sib analyses only work with one or two generations (Falconer, 1989). Moreover, they cannot take into account all relationships in a complex pedigree. In our dataset, for example, some individuals shared one pair of their grandparents and others one or both of their parents. Maximum likelihood analyses are more flexible and suited for analysing such datasets.

We used the nf3-program Quercus of R.G. Shaw & F.H. Shaw to perform REML analyses (Shaw, 1987). The program was adapted by one of us (J.J.W.) for use on a personal computer under Turbo Pascal (version 7.0) and for analysing three generations. Females in Generation 1 were assumed to have mated only once in the wild (see above). Under REML, relationship matrices are used that contain the expected covariances between all individuals. For example, additive genetic covariances are 0.5 between parents and offspring, 0.25 for half sibs and 1 within individuals, for dominance variance 0.25, 0 and 1, respectively, for residual variance 0, 0 and 1, respectively (Roff, 1997; Lynch & Walsh, 1998). These matrices are used to calculate the likelihood for a certain set of (co)variance components. In an iterative procedure these calculations are repeated until the set of components that result in the maximum likelihood (ML) is found.

REML analyses can be constrained, for example to only positive values for variances, or to a fixed value of 0 for dominance variance. The significance of a variance component can be tested by comparing the ML under a full model with the ML under a model in which it is constrained to 0. Twice the difference between the MLs follows a χ^2 distribution with one degree of freedom. REML analyses can also be used to compare different models of genetic determination, such as autosomal and sex-linked inheritance. Under sex-linked inheritance a matrix describing the genetic covariances under sex linkage between individuals is added to the model. For example, when the heterogametic sex is the female, as is the case in butterflies, the expected covariance between a grandmother and granddaughter is 0.25 for additive effects for both maternal and paternal grandmothers. For sex-linked effects, however, the expected covariance is 0 for maternal grandmother–granddaughter and 0.5 for paternal grandmother–granddaughter (table 24.1 in Lynch & Walsh, 1998).

Several models were examined with REML analysis, to evaluate the genetics of oviposition preference in *P. c-album*. The first model assumed no sex linkage and evaluated the genetic variance (V_G) and the residual variance (V_R) component. The second model assumed only variance due to sex linkage and sex-dependent gene expression and thus evaluated V_{XY} and V_R. The third model analysed V_G, V_{XY} and V_R. Finally, a model was analysed in which the genetic variance was split over an additive variance component (V_A) and a dominance variance component (V_D). Because the software program could not fit a model with four variance components, no model was analysed with V_A, V_D, V_{XY} and V_R. In any case the dataset was probably too small for the evaluation of such a large model. Preference differed between the generations; therefore generation was added as a fixed effect to the model. Doing this does not, however, completely remove the generation effect. If there is a genotype by generation effect (e.g. some genotypes show larger differences between generations than others) this will influence the estimated variance components. Thus the three models were also run on a dataset containing only oviposition preferences for the third generation.

STUDY 2

For this investigation, females were collected at four localities over a larger area than in Study 1: Järna, south-west of Stockholm ($N = 2$ females with mated female offspring represented in F_1); Järva, close to the city limits of Stockholm ($N = 6$ females); Åkersberga, north-east of Stockholm ($N = 7$ females); Norrtälje, further to the north-east ($N = 2$ females). The approximate distance between the two most distant sites (Järna and Norrtälje) is 110 km.

Preferences in the parental generation were determined as in Study 1. From this generation 30 hatchlings from each female were reared individually as in Study 1 to produce the F_1 generation. We aimed to investigate the preferences of three F_1 females for each of the 17 parental females (dames) but the actual figure varied from one to five, depending on the availability of females that mated in the cages. The most extreme individuals in the F_1 generation (independent of origin) were used to initiate two selection lines (*Urtica*-preference and *Salix*-preference, relative to the other host plant) and a control group. There were 69 mated females in F_1, of which eight were assigned to each selected line. Offspring of eight females were randomly assigned to the control line, independently of whether these females were also founders of a selected line.

Ten offspring from each founder female (or 20 if they were assigned to two selection lines) were reared individually to produce generation F_2, with all adults in a given line being put together to mate in a larger cage. In the following generations, we allowed the 30–50% most extreme females in the selected lines to be represented by offspring in the next generation; the percentage varied slightly depending on the total number of mated females available and on the distribution of preferences over females. In the control line, we randomly assigned a similar number of females to be represented in the next generation. Fifteen to 20 hatchlings from each female were reared individually in this and the following generations, and the resulting adults mated in cages, as above. This experiment ended in 2001 with generation F_6, after two instances of winter hibernation in the adult stage (in generations F_3 and F_6). Selection was performed on four occasions between generations F_1 and F_5.

In generation F_5 we allowed matings to take place both within and between the two selected lines, in a search for sex-linked inheritance of the difference in preference between them. Hence, only offspring of females that mated with males belonging to their own selection line are shown as F_6 in Figures 5 and 6. Due to the resulting lower number of individuals remaining in the pure selected lines, and due to little variation in preference in this late summer generation, no further selection was performed.

RESULTS

STUDY 1

Preferences in the parental generation

As Janz *et al*. (1994) found, there was considerable variation among females in the relative preferences for *U. dioica* in the parental generation (Fig. 1; $P < 0.05$ in either a parametric ANOVA or a non-parametric Kruskal–Wallis ANOVA). However, we found that all females oviposited more (typically 60–80% of the eggs) on *U. dioica* than they did on *S. caprea* on average (Fig. 1).

Non-genetic causes of variation may have influenced these patterns. One indication that this was so was the presence of temporal trends in the data. Females laid an increasing proportion of eggs on *U. dioica* over time (regression against day since first oviposition, $r^2 = 0.106$, $P < 0.01$, $N = 66$, all females pooled). This could indicate a process of learning, with females responding to the presence of a favoured host in the cages. However, an even stronger temporal trend (in terms of explanatory power) was the one found between day of capture and the average proportion of eggs laid on *U. dioca.* This proportion was higher for females captured late in the season (regression, $r^2 = 0.565$, $P < 0.05$, $N = 9$ females).

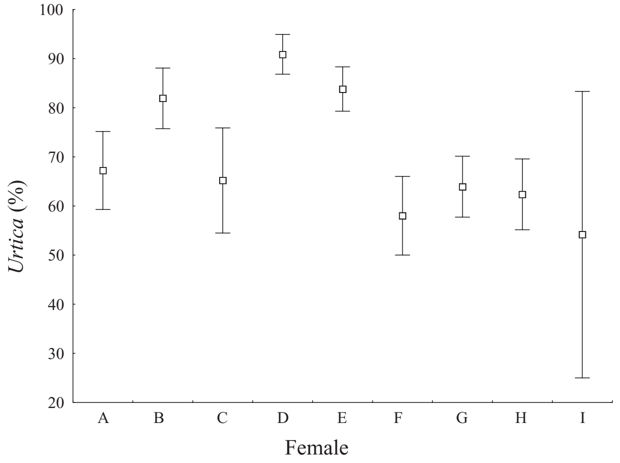

Figure 1. Host-plant preferences in *Polygonia c-album* (percentage of eggs laid on *Urtica* daily): means and SE (variation among days) of individual females caught in the wild (parental generation of Study 1) when given a choice between *U. dioica* and *Salix caprea* in flight cages.

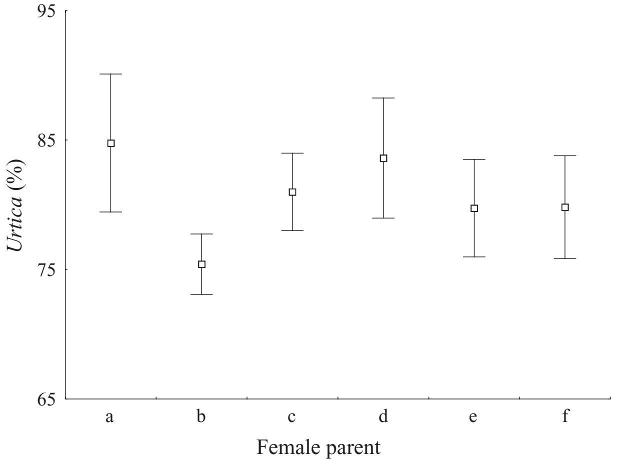

Figure 2. Host-plant preferences in *Polygonia c-album* (percentage of eggs laid on *Urtica*) of females of the first laboratory-reared generation (F_1) when given a choice between *U. dioica* and *Salix caprea* in flight cages. Means and SE (variation among females) for six families. Letters correspond to parental females in Fig. 1.

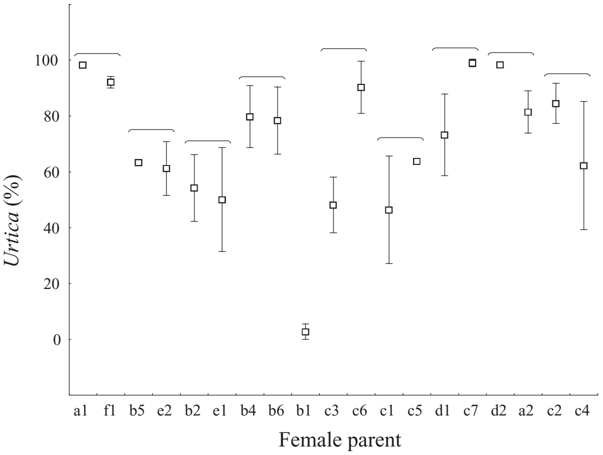

Figure 3. Host-plant preferences in *Polygonia c-album* (percentage of eggs laid on *Urtica*) of females of the second laboratory generation (F_2) when given a choice between *U. dioica* and *Salix caprea* in flight cages. Data points show means and SE (variation among females) for full sibships whereas brackets join pairs of half sibships, sharing a male parent. One family (b1) lacks half sibs. Letters correspond to parental females in Fig. 2.

Preferences in the F_1 generation

The oviposition preferences of 35 females were scored in the F_1 generation. They were the progeny of six of the parental females, the remaining females having failed to produce any female offspring that mated in the cages. As found earlier (Nylin, 1988; Janz, 1998), the proportion of eggs laid on *U. dioca* was higher in this summer form generation, typically 70–90% (Fig. 2).

Again, variation among females was pronounced (ANOVA, $P < 0.01$, $N = 35$ females), despite the fact that all females were now naive. There was no significant effect of the parental female on this variation (Fig. 2; parametric ANOVA or Kruskal–Wallis ANOVA, not significant) and regression between preferences in the parental generation and averages from all their offspring in the F_1 generation suggested no additive genetic variation ($r^2 = 0.005$, not significant, $N = 6$). However, no actual estimate of heritability was attempted, since this result was confounded not only by a small sample size in the parental generation but also by the possibility of induced preferences (learning) in the field prior to capture.

The temporal trend in the data persisted, albeit with low explanatory power (regression against day of experiment, $r^2 = 0.06$, $P < 0.001$, $N = 35$, pooling all females). Thus, the variation among females was not strongly influenced by this trend.

Preferences in the F_2 generation

The preferences of 70 females were scored in the F_2 generation (Fig. 3), and the variation among females was even greater than it was in the preceding generations. Averages for individual females ranged from 0% to 100% (ANOVA, $P < 0.001$, $N = 70$; Kruskal–Wallis ANOVA, $P < 0.05$–$P < 0.001$ with an arbitrary sample of ten females). Preferences for *Urtica* were somewhat lower than they were in the F_1 generation (typically 50–80%). This in all probability resulted from the procedure, adopted in this generation, of placing *Salix* plants higher in the cage compared with *Urtica* plants. Individual females searching for hosts only high in the cages, closer to the source of light, would have oviposited predominantly on *Salix.* This extra source of variation may have led us to underestimate the genetic component of host-plant preference, unless the actual search strategies of females are genetically tightly linked to their preferences (a clear possibility in this species, which feeds on plants of all different growth forms, from herbs to trees).

A temporal trend was again present but did not influence the data strongly, as the explanatory power was low (regression against day of experiment, $r^2 = 0.03$, $P < 0.01$, $N = 70$).

Ovipositing females of this generation were related to each other in a known but complex manner; they were a mixture of full sibs and half sibs, as explained above, and also shared grandparents on one or both parental sides. Effects of lineage are analysed in more detail in the next section, but some patterns can be mentioned here (based on average preferences of individual females). Shared female or male parents both had significant effects on host-plant preferences, ignoring other relationships (Fig. 3; dams: ANOVA, $P < 0.05$, $N = 19$; sires: ANOVA, $P < 0.001$, $N = 10$; Kruskal–Wallis ANOVA, $P < 0.01$, $N = 10$). Family b1, which lacked half sibs, were omitted from all statistical analyses in this section.

In Figure 3, half-sib females are grouped together, according to the design that was intended to reveal whether there was paternal inheritance of host-plant choice (as expected under X-linkage). It can be seen that pairs of half sibships (sharing male but not female parent) in some cases really did show similar preferences (corresponding to the significant effect of 'sires' above), but in other cases they differed strongly (due to the independent effect of 'dams'). When effects of both sires and dams were instead analysed together (GLM nested ANOVA, dams nested within sires) none of these effects were significant when they were treated as random factors, the strictly correct procedure. However, since there were only two dams per sire from which to estimate the variance, this result is statistically unreliable. Sires explained about twice as much of the variance as did dams (13.9% and 7.2%, respectively), but since neither effect was significant this result must be interpreted very cautiously.

The effects of shared grandparents were significant on both parental sides, despite the low sample sizes (mothers to sires, ANOVA, $P < 0.05$, Kruskal–Wallis ANOVA, $P < 0.05$, $N = 4$; mothers to dams, ANOVA, $P < 0.05$, Kruskal–Wallis ANOVA, $P < 0.05$, $N = 6$).

A regression of female preferences in F_1 against those shown by their offspring in F_2 was significant ($r^2 = 0.23$, $P < 0.05$, $N = 19$).

Taken together, these patterns of similar preferences in females that shared parents and/or grandparents suggest a strong genetic component of the variation in F_2, but not a major role of genes on the X chromosome. If preferences were strongly sex-linked and carried on the X chromosome we would not expect significant effects of shared mothers to dams (i.e. genes from females caught in the wild) to remain in the F_2 generation.

Genetic analysis by REML

So far we have largely ignored the complex breeding design and the known pedigrees for mating individuals. To take these into account, we performed REML analyses.

The models incorporating sex-linked inheritance did not provide a better fit than did models without sex linkage (Table 1). On the contrary, the model without sex-linked inheritance had a significantly better fit in the three-generation analysis. Genetic variance was high and in the one-generation analysis significantly larger than 0, while the sex-linked variance component was never significantly different from 0. A model with both a genetic variance component and a sex-linked variance component did not converge in the three-generation analysis, and gave a negative sex-linked component in the one-generation analysis. Splitting the genetic variance into an additive and a

Table 1. Variance components in Study 1 estimated by residual error maximum likelihood (REML) analysis under three models

Model	V_G	V_{XY}	V_R	V_P	ML
Three generations					
V_G	0.221		–0.077	0.145	60.1656
%	152.9		–52.9		**0.0057**
P	0.082				
V_{XY}		0.056	0.095	0.151	56.3341
%		36.9	63.1		
P		0.570			
$V_G + V_{XY}$	Not converged				
Only 3rd generation					
V_G	0.240		–0.081	0.259	26.8727
%	131.2		–31.2		0.319
P	**0.006**				
V_{XY}		0.308	–0.062	0.248	26.3731
%		126.6	–26.6		
P		0.155			
$V_G + V_{XY}$	0.305	–0.113	0.054	0.246	25.3144
%	123.3%	–45.4	22.1		
P	0.146	–			

Numbers in bold type are statistically significant probabilities.

dominance component did not improve the models and gave mixed results. In the three-generation analysis all variance ended up in the dominance variance component whereas in the one-generation analysis it was split evenly over the dominance and additive components. This difference might be due to a generation by genotype interaction (see Methods). Dominance and additive variance components on their own were never significantly different from 0. Thus there was significant genetic variance for oviposition preference, but to what extent it was due to additive and/or dominance effects is not clear, and there was no evidence of sex-linked inheritance.

Sib matings

There was no tendency for sib matings to be avoided in cages with two families present. In fact, sib matings were more common than were matings between unrelated individuals, but this seemed to be due mostly to differences in viability among families, sometimes leaving individuals with mostly sibs to mate with. Because of this complication, we did not perform a statistical analysis of mating patterns. The mortality was higher in families of offspring resulting from sib matings, on average 41% compared with 11% for non-sib matings ($P < 0.05$, $\chi^2 = 33.67$). The excess mortality was due almost entirely to larvae failing to moult properly, with newly moulted larvae leaking body fluids.

STUDY 2

In the parental generation of wild-caught females there was significant variation among localities (not shown; GLM ANOVA, female random effect nested in localities, $P < 0.05$, $N = 4$ localities, $N = 17$ females) but not independently among females ($P = 0.50$). The significant effect of locality disappeared if the two females from the most southern population (Järna), with high preference for *Urtica*, were removed from the analysis.

Results from these wild-caught females were associated with problems such as the possibility of learning in the field and variation in day of collection (in view of the seasonal trend noted above). In the F_1 offspring of these females there was no significant effect of locality (Fig. 4; GLM ANOVA, female random effect nested in localities, $P = 0.83$, $N = 4$ localities, $N = 66$ females), and the variation among females was instead significant ($P < 0.01$). If day of oviposition was added to the model as a continuous factor (covariate) this effect was also significant ($P < 0.05$) due to a temporal pattern of increasing preference for *Urtica* noted also in Study 1. Effect of locality remained not significant ($P = 0.66$) and effect of female was highly significant ($P < 0.001$).

In the selection experiment that followed, the selection lines diverged immediately (but not to significantly different preferences) in generation F_2, and the *Urtica* line remained above the *Salix* line in preference in the following generations (Fig. 5). Averages in Figures 5 and 6 are shown with the control line subtracted in order to control for the large seasonal variation. A greater difference between the selected lines ($P < 0.03$ in a one-tailed *t*-test) could be seen in generation F_3, after the first hibernation. In this spring form

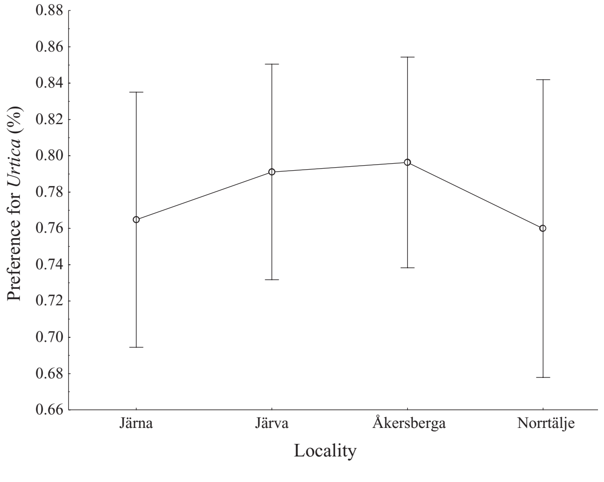

Figure 4. Host-plant preferences in *Polygonia c-album* (percentage of eggs laid on *Urtica*) of females of the first laboratory-reared generation (F_1) in Study 2 when given a choice between *U. dioica* and *Salix caprea* in flight cages. Means and SE (variation among females) for females originating from four localities in the Stockholm area, arranged from south to north.

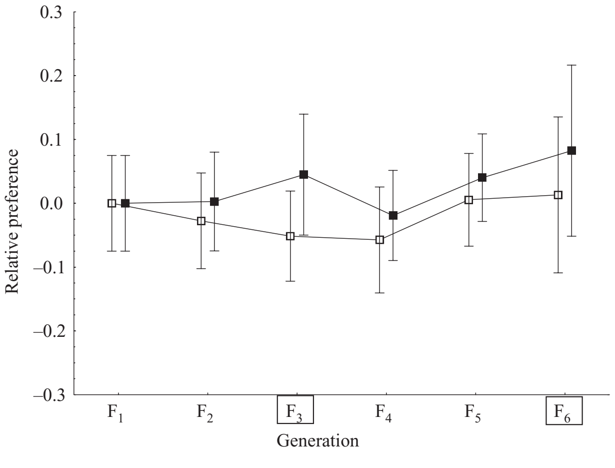

Figure 5. Host-plant preference averages (relative to the control line) in *Polygonia c-album* for *Urtica dioica* when given a choice with *Salix caprea* in the two selected lines over the span of the selection experiment (■ = *Urtica* line; □ = *Salix* line). Bars show SE for variation among females. In F_1 both averages are for the combined selected lines. Generations in squares are spring form adults.

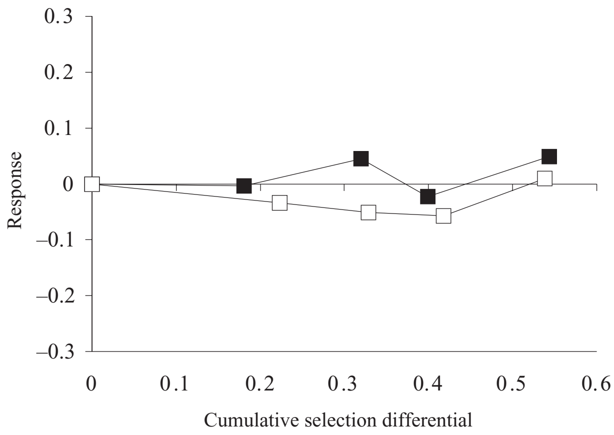

Figure 6. Host-plant preference averages (relative to the control line) in *Polygonia c-album* for *Urtica dioica* when given a choice with *Salix caprea* in the two selected lines as a function of the cumulative selection differential. ■ = *Urtica* line; ☐ = *Salix* line.

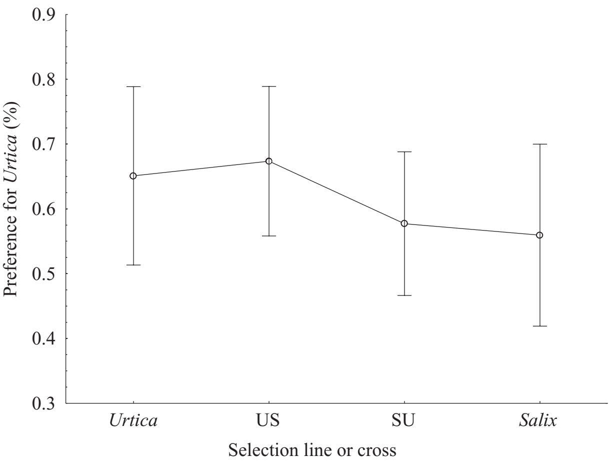

Figure 7. Host-plant preference in *Polygonia c-album* for *Urtica dioica* relative to *Salix caprea* (averages and SE among females) in the two selected lines in F_6 as well as for the offspring of crosses between the lines. For the crosses, maternal genotype is indicated by U = *Urtica* line or S = *Salix* line followed by the paternal genotype.

generation, females were less specialized on *Urtica* and there was greater scope for genetic variation to be expressed. The selection lines did not diverge further over the span of the experiment (Fig. 5), and indeed there was very little response to selection after generation F_3, suggesting a very small additive component of the genetic variation and low narrow-sense heritability (Fig. 6).

When the *Urtica* and *Salix* selection lines were allowed to cross to produce the F_6 generation, there were no significant differences in preferences among the four resulting categories: pure stock from the two selected lines and the two reciprocal crosses between them (Fig. 7). However, rather than a pattern of females from crosses resembling their paternal genotype, as would be expected with X-linkage, they tended to resemble their maternal genotype (Fig. 7).

DISCUSSION

The patterns reported here support the prediction that considerable genetic variation in host-plant preference can exist even within a single local population of *Polygonia c-album* (Study 1), despite the strong fitness effects of larval host plants. Presumably this is a result of fitness trade-offs among hosts (see Krainacker *et al*., 1987 for a similar explanation applied to the Mediterranean fruit fly), acting in combination with relaxed selection for using 'fast' hosts (Nylin, 1988; Scriber & Lederhouse, 1992; Janz *et al*., 1994; Janz, 1998). In the initial phase of Study 2 we found strong variation among females, but not among the source localities for these females, suggesting an open population structure with strong gene flow. The high percentage of inviable offspring from sib matings in Study 1 is consistent with such high levels of gene flow, with little purging of lethal recessive alleles by local inbreeding. The lack of a cumulative response to selection for preference for *Urtica* or *Salix* (Study 2), as well as the lack of an intermediate trait value when the selections lines were finally crossed, demonstrates that there is little additive genetic variance. Instead, most of the observed variation among females may be due to major genes with dominance effects. Sex-linked inheritance does not seem to play a major role on this local scale.

One important consequence of the X-linked inheritance of preference differences between Swedish and English *P. c-album* found by Janz (1998), with expression of the paternal genotype rather than an intermediate trait value, is that it suggests that differences in host specificity can be determined by major genes. According to the influential model for host selection proposed by Courtney *et al*. (1989), host ranking and host specificity are different traits, with host specificity determined by a threshold value for acceptance that varies with the motivational state of the insect. Such behavioural, quantitative traits have been viewed as most likely to be polygenic (cf. Ritchie & Phillips, 1998) and best understood in terms of quantitative genetics (Courtney *et al*., 1989). Polygenic inheritance would present a problem for applying theory on a link between sex linkage and speciation (Prowell, 1998), since such theory models effects of major genes situated on the sex chromosomes (Ritchie & Phillips, 1998). The results of our study further support, also on a more local scale, that variation in host specificity may in fact be determined by major genes. This broadens the scope of theory on sex linkage and speciation in the Lepidoptera [and perhaps other organisms (Prowell, 1998)] to include also female host specificity. Being a sex-limited component of habitat choice, this trait has the clear potential to be involved in premating isolation (Prowell, 1998; Ritchie & Phillips, 1998).

We observed some presumably non-genetic patterns in this study. A temporal trend with an increasing percentage of eggs laid on *Urtica* was seen in all experiments. This could suggest the presence of learning, but if so learning explained very little of the variation in preference, compared with lineage effects (genetics). Another possible explanation for the temporal trend could be a life-history pattern. Early oviposition could take place more randomly among potential hosts (to ensure that at least some eggs will be laid on hosts), whereas later oviposition could be more discriminating in favour of the preferred hosts (cf. Larsson & Ekbom, 1995). The seasonal trend among the females caught in the wild is also consistent with this explanation. It should be added that we also observed large within-individual variation that did not conform to this temporal trend, with some females switching preference from day to day, and that few females (especially in the spring generation) laid all their eggs on either *U. dioica* or *S. caprea* (unpubl. observ.). This may suggest an element of risk-spreading or bet-hedging as an additional factor behind the strong polyphagy displayed by *P. c-album*, and studies to test this hypothesis are under way.

Finally, the results reported here have consequences for the ongoing discussion regarding how insect host-plant ranges are determined and how easily and frequently they change as a result of evolution. Similar host-plant rankings over large areas, seen in *P. c-album* as well as in, for example, *P. zeliacon* (Wehling & Thompson (1997), suggest evolutionary conservatism of host-plant preferences. Other insects instead show rapid changes in host-plant preference (e.g. Singer, Thomas & Parmesan, 1993). However, this seeming contradiction may be deceptive. As can be seen from our study, average preferences in a local population may hide substantial local variation among individuals, which (if some part of this variation is genetic) means that the population is ready to respond to changing selection pressures. Hence, evolutionary conservatism may not equal strongly constrained evolution. Similar preference patterns over large areas may simply reflect similar selection, i.e. the ‘best host plants’ are the best (when all fitness effects are summed) everywhere. In addition, with an open population structure gene flow can eliminate local adaptation but preserve genetic variability, especially when the additive component of variation is small and hence the response to selection is slow. As noted above, this seems to be the situation in *P. c-album*. Apparently, considerable genetic variation for host-plant preferences may be present in local populations without changes in the average ranking of host plants.

CONCLUSIONS

We found no evidence of X-linkage of host preference within the Stockholm population, in contrast to the situation when populations of *P. c-album* from Sweden and England are compared (Janz, 1998), and in contrast to differences between several pairs of closely related species of butterfly and other Lepidoptera (Sperling, 1994; Prowell, 1998). This could suggest that such sex-linked genetic differences arise on a larger geographical and evolutionary scale, but still within species (i.e. preceding and potentially aiding speciation). Future studies involving progressively more distant populations of *P. c-album* will address the question regarding at what scale sex-linkage becomes apparent. Judging from the combined results, it may be that a few major genes influence host-plant preference in the highly polyphagous Stockholm population of *P. c-album*. At least some of them are evidently autosomal, but others may be carried on the sex chromosomes (X or Y). The X-linked genetic difference in preference between populations of this species (Janz, 1998) could be the result of a more or less random fixation of one of these genes at some point in the history of the populations. However, the pattern that differences between closely related Lepidopteran species are X-linked disproportionately often seems to suggest that such genetic differences are somehow involved in speciation in this group of insects. Thus, it may be the case that variation at lower levels have varying genetic backgrounds, but differences among individuals caused by X-linked genes have a higher probability of being carried over to the level of species differences.

ACKNOWLEDGEMENTS

This study was financially supported by grants from the Swedish Research Council (V.R.) to S.N. and N.J. Discussions with Mike Singer inspired parts of the Discussion.

REFERENCES

Beck SD. 1980. *Insect photoperiodism*. New York and London: Academic Press.

Bernays EA, Chapman RF. 1994. *Host–plant selection by phytophagous insects*. London: Chapman & Hall.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist* **130:** 113–146.

Courtney SP, Chen GK, Gardner A. 1989. A general model for individual host selection. *Oikos* **55:** 55–65.

Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study in coevolution. *Evolution* **18:** 586–608.

Ellers J, Boggs CL. 2002. The evolution of wing colour in *Colias* butterflies: heritability, sex linkage, and population divergence. *Evolution* **56:** 836–840.

Falconer DS. 1989. *Introduction to quantitative genetics*. Harlow: Longman.

Hawthorne DJ, Via S. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* **412:** 904–907.

Jaenike J. 1989. Genetics of butterfly–hostplant associations. *Trends in Ecology and Evolution* **4:** 34–35.

Janz N. 1998. Sex-linked inheritance of host-plant specialization in a polyphagous butterfly. *Proceedings of the Royal Society of London B* **265:** 1–4.

Janz N. 2003. Sex linkage of host plant use in butterflies. In: Boggs CL, Watt WB, Ehrlich PR, eds. *Butterflies: ecology and evolution taking flight*. Chigaco: Chicago University Press, 229–239.

Janz N, Nylin S. 1997. The role of female search behaviour in determining host plant range in plant feeding insects: a test of the information processing hypothesis. *Proceedings of the Royal Society of London B* **264:** 701–707.

Janz N, Nylin S, Wedell N. 1994. Host plant utilization in the comma butterfly: sources of variation and evolutionary implications. *Oecologia* **99:** 132–140.

Krainacker DA, Carey JR, Vargas RI. 1987. Effect of larval host on life history traits of the Mediterranean fruit fly, *Ceratitis capitata*. *Oecologia* **73:** 583–590.

Larsson S, Ekbom B. 1995. Oviposition mistakes in herbivorous insects: confusion or a step towards a new host plant? *Oikos* **72:** 155–160.

Lynch M, Walsh B. 1998. *Genetics and analysis of quantitative traits*. New York: Chapman & Hall.

Mitter C, Farrel B, Wiegmann B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist* **132:** 107–128.

Ng D. 1988. A novel level of interactions in plant-insect systems. *Nature* **334:** 611–613.

Nylin S. 1988. Host plant specialization and seasonality in a polyphagous butterfly, *Polygonia c-album* (Nymphalidae). *Oikos* **53:** 381–386.

Nylin S. 1989. Effects of changing photoperiods in the life cycle regulation of the comma butterfly, *Polygonia c-album* (Nymphalidae). *Ecological Entomology* **14:** 209–218.

Nylin S. 1992. Seasonal plasticity in life history traits: growth and development in *Polygonia c-album* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **47:** 301–323.

Nylin S, Janz N. 1993. Oviposition preference and larval performance in *Polygonia c-album* (Lepidoptera: Nymphalidae) – the choice between bad and worse. *Ecological Entomology* **18:** 394–398.

Nylin S, Janz N. 1996. Host plant preferences in the comma butterfly (*Polygonia c-album*): do parents and offspring agree? *Ecoscience* **3:** 285–289.

Nylin S, Janz N, Wedell N. 1996. Oviposition plant preference and offspring performance in the comma butterfly: correlations and conflicts. *Entomologia Experimentalis et Applicata* **80:** 141–144.

Prowell DP. 1998. Sex linkage and speciation in Lepidoptera. In: Howard DJ, Berlocher SH, eds. *Endless forms: species and speciation*. Oxford: Oxford University Press, 309–319.

Ritchie MG, Phillips SDF. 1998. The genetics of sexual isolation. In: Howard DJ, Berlocher SH, eds. *Endless forms: species and speciation*. Oxford: Oxford University Press, 291–308.

Roff DA. 1997. *Evolutionary quantitative genetics*. New York: Chapman & Hall.

Scriber JM. 1994. Climatic legacies and sex chromosomes: latitudinal patterns of voltinism, diapause, body size, and host-plant selection on two species of swallowtail butterflies at their hybrid zone. In: Danks HV, ed. *Insect life-cycle polymorphism: theory, evolution and ecological consequences for seasonality and diapause control.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Scriber JM, Giebink BL, Snider D. 1991. Reciprocal latitudinal clines in oviposition behaviour of *Papilio glaucus* and *P. canadensis* across the Great Lakes hybrid zone: possible sex-linkage of oviposition preferences. *Oecologia* **87:** 360–368.

Scriber JM, Lederhouse RC. 1992. The thermal environment as a resource dictating geographic patterns of feeding specialization of insect herbivores. In: Hunter MR, Ohgushi T, Price PW, eds. *Effect of resource distribution on animal–plant interactions*. New York: Academic Press, 429–466.

Shaw RG. 1987. Maximum-Likelihood approaches applied to quantitative genetics of natural populations. *Evolution* **41:** 812–826.

Singer MC, Thomas CD, Parmesan C. 1993. Rapid human-induced evolution of insect–host associations. *Nature* **366:** 681–683.

Sperling FAH. 1994. Sex-linked genes and species differences in Lepidoptera. *Canadian Entomologist* **126:** 807–818.

Thompson JN. 1988. Evolutionary genetics of oviposition preference in swallowtail butterflies. *Evolution* **42:** 1223–1234.

Thompson JN, Pellmyr O. 1991. Evolution of oviposition behavior and host preference in Lepidoptera. *Annual Review of Entomology* **36:** 65–89.

Via S. 1990. Ecological genetics and host adaptation in herbivorous insects: the experimental study of evolution in natural and agricultural systems. *Annual Review of Entomology* **35:** 421–446.

Wedell N, Nylin S, Janz N. 1997. Effects of larval host plant and sex on the propensity to enter diapause in the comma butterfly. *Oikos* **78:** 569–575.

Wehling WF, Thompson JN. 1997. Evolutionary conservatism of oviposition preference in a widespread polyphagous insect herbivore, *Papilio zelicaon*. *Oecologia* **111:** 209–215.