A continuum of genetic divergence from sympatric host races to species in the pea aphid complex

Jean Peccoud[a], Anthony Ollivier[a], Manuel Plantegenest[b], and Jean-Christophe Simon[a,1]

[a]Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1099 BiO3P, Domaine de la Motte, 35653 Le Rheu, France; and [b]Agrocampus Ouest, Unité Mixte de Recherche 1099 BiO3P, 65 rue de St. Brieuc, 35042 Rennes, France

Edited by Francisco J. Ayala, University of California, Irvine, CA, and approved March 13, 2009 (received for review November 3, 2008)

Sympatric populations of insects adapted to different host plants, i.e., host races, are good models to investigate how natural selection can promote speciation in the face of ongoing gene flow. However, host races are documented in very few model systems and their gradual evolution into good species, as assumed under a Darwinian view of species formation, lacks strong empirical support. We aim at resolving this uncertainty by investigating host specialization and gene flow among populations of the pea aphid complex, *Acyrthosiphon pisum*. Genetic markers and tests of host plant specificity indicate the existence of at least 11 well-distinguished sympatric populations associated with different host plants in Western Europe. Population assignment tests show variable migration and hybridization rates among sympatric populations, delineating 8 host races and 3 possible species. Notably, hybridization correlates negatively with genetic differentiation, forming a continuum of population divergence toward virtually complete speciation. The pea aphid complex thus illustrates how ecological divergence can be sustained among many hybridizing populations and how insect host races blend into species by gradual reduction of gene flow.

divergent selection | ecological speciation | hybridization | sympatric speciation | phytophagous insects

The beginning of this century has seen the reconciliation of Darwin's view that species gradually emerge by the primary action of natural selection, with the modern evolutionary synthesis defining speciation as the evolution of reproductive isolation (1–4). Adaptations to different habitats and resources may induce reproductive isolation (reviewed in refs. 1, 5, and 6) if morphology or niche selection determines mate choice and if hybrids between ecologically divergent taxa are unfit in the parental environments. Theoretical models and recent empirical studies (reviewed in refs. 7 and 8) have shown that these ecological reproductive barriers may evolve without the geographical separation of populations, an extrinsic barrier to hybridization (gene flow) that was long thought to be a requisite for speciation in animals (2). Divergence in sympatry (i.e., within the same region) implies a more gradual reduction of gene flow that should reflect increasing reproductive isolation and the continuum of speciation from polymorphic populations to good species. Such evolution is not observable in nature at reasonable timescales, but it could be inferred from the multiple stages of sympatric divergence observed within a variety of organisms (9). In phytophagous insects, "host races" constitute these intermediate stages of speciation. Host races refer to sympatric populations in partial reproductive isolation that are specialized to different host plants (10–12). Although host race formation can be rapid (12–14), their transition into full species and the reduction of gene flow during the final stages of speciation have remained difficult to trace. Host races are characterized in very few insect species (11), and individual case studies reveal only an isolated snapshot of what is an ongoing process.

Here, we illustrate how host races may evolve into species, by documenting a continuum of population differentiation toward virtually complete speciation in sympatry within the same biological model, the pea aphid, *Acyrthosiphon pisum* Harris (Homoptera: Aphididae). This aphid thrives on dozens of plant genera of the legume family (Fabaceae) across its palearctic native range and on a few major legume crops introduced into other continents (15). Although named as a single species, this complex encompasses sympatric populations showing differential preference and fitness on specific host plants (16–20); hereafter, we refer to such populations as "biotypes," a word that is useful for cases in which genetic bases of variation have not yet been established. Among these morphologically similar populations, only those feeding on a few crops have been characterized genetically. In North America, populations specialized on red clover and alfalfa are genetically distinct (21), and genetic trade-offs prevent the optimal use of both host species by pea aphids lineages (22). In Europe, 2 host-adapted populations coexist on the same plant species with another one specialized on pea and broad bean, which exhibits a higher genetic divergence (23, 24). The precise levels of reproductive isolation between these host-specialized populations remain nonetheless unresolved. Other populations feeding on wild plants in Western Europe have been recognized at the rank of subspecies or species on the basis of subjective morphological or ecological criteria (15, 25). The pea aphid constitutes therefore an ideal model for studying a range of stages of speciation in sympatry.

Results and Discussion

Multiple Host-Specialized Sympatric Populations. We examined western European pea aphids on 19 widely distributed, sympatric host species, mostly wild plants, using microsatellite DNA markers and laboratory tests of host plant specificity. The Bayesian assignment method implemented in the program Structure (26–28), with no a priori information on the origin of individuals, identified 11 well-distinguished genetic clusters, or populations, associated with different host plants and not with sampling locations (Fig. 1*A*). This sharp population genetic structure is mirrored by clear differences in performance (survival and growth) of aphid lineages tested on representative plants in the laboratory (Fig. 1*B*). Consistent with expectations, pea aphid lineages performed better on the test plant from which individuals of their genetic population were usually collected [supporting information (*SI*) Fig. S1]. Phenotypic data thus demonstrate that genetic clusters correspond to 11 host-specialized populations, which we hereafter call biotypes.

Three of these biotypes, noted "E" on clovers (*Trifolium*), "G"

Author contributions: J.P., M.P., and J.-C.S. designed research; J.P. and A.O. performed research; J.P. contributed new reagents/analytic tools; J.P. analyzed data; and J.P., M.P., and J.-C.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The DNA sequences reported in this paper have been deposited in the GenBank database (accession nos. FJ855137 to FJ855200).

[1]To whom correspondence should be addressed. E-mail: jean-christophe.simon@rennes.inra.fr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0811117106/DCSupplemental.

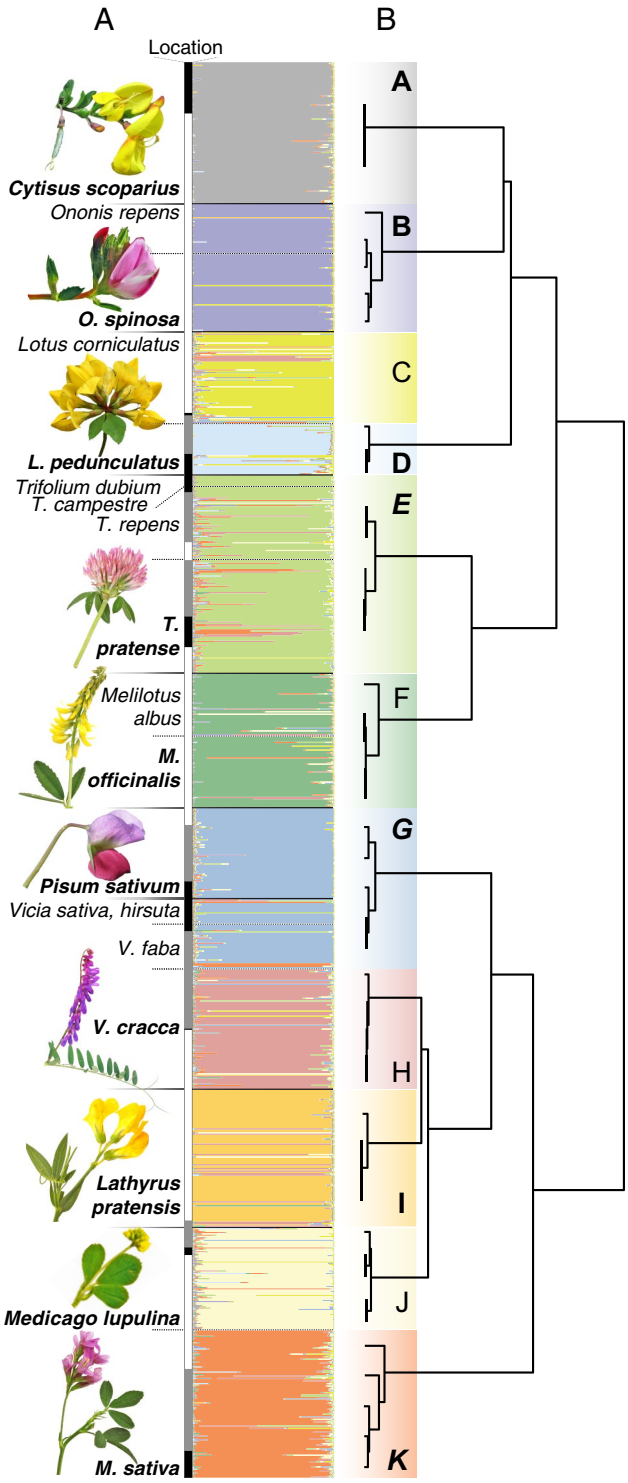

Fig. 1. Genetic structure and host specialization in western European pea aphids. (*A*) Population structure: horizontal segments represent 1,090 individuals (14-microsatellite genotypes), and colors represent their proportions of ancestry to inferred populations coded from A to K (see *B*). Genotypes are sorted by host plants and then by locations, as indicated by the vertical bar on the left. White, eastern France; black, northwestern France; gray, eastern Germany. (*B*) A phenogram of 45 aphid laboratory lineages assigned to 10 populations and based on their performance (survival and growth) on 10 representative plant species (photographs and names in boldface type in *A*). Branch lengths thus represent differences in host specialization on these plants. Italicized letters indicate biotypes that were previously characterized genetically (23, 24) and letters in boldface type indicate biotypes whose existence was suspected from earlier host specialization tests (17, 18).

on pea (*Pisum*) and annual vetches (*Vicia faba*, *hirsuta*, and *sativa*), and "K" on alfalfa (*Medicago sativa*), correspond to populations that were previously characterized genetically on narrower host ranges (23, 24). The existence of biotypes "A" on *Cytisus scoparius*, "B" on *Ononis* sp., "D" on *Lotus pedunculatus*, and "I" on *Lathyrus pratensis* was expected, given earlier phenotypic measures of host specialization on these plants (17, 18). The 4 remaining biotypes "C, H, J, and F" were not identified before the present study, to our knowledge. Notably, the first 3 are specialized on plant species belonging to genera harboring other biotypes: *L. corniculatus*, *V. cracca*, and *M. lupulina*, respectively, with biotype F being specialized on 2 species of *Melilotus.* Given that many other legume genera harbor pea aphids (25), the discovery of these 4 biotypes strongly suggests the existence of many other biotypes in the pea aphid complex.

Gene Flow and Genetic Differentiation Among Pea Aphid Biotypes. Because pea aphids grow and mate on their host plants (25, 29), the association between biotype and plant species informs us on the potential for gene flow. The Bayesian method implemented in the program Geneclass (30) revealed the presence of "migrant" aphids, which correspond to individuals collected on a plant species that is not typical of their biotype (Table 1). Host specialization tests on laboratory lineages initiated by some of these migrant individuals were consistent with their genetic assignment, as they showed similar patterns of host performance to other lineages of the same genetic cluster (Fig. S1). Immigration rates are averaged at 9.4% across host species and indicate potential gene flow among biotypes.

To quantify the levels of actual gene flow, we looked for first-generation hybrids, i.e., individuals sharing ancestries with 2 biotypes, using the programs Structure and NewHybrids (31). We also measured the genetic differentiation at microsatellite loci between biotypes and in sympatry, using hierarchical analyses of molecular variance (32, 33). The frequency of hybrids per biotype pair appeared moderate (Fig. 2, circles). However, because of the various parental origins for hybrids within this complex of 11 biotypes (Table S1), their cumulative proportion per biotype could reach 9% (Fig. 2, diamonds) and negatively correlates with their level of genetic differentiation. This correlation can be explained if the proportion of hybrids partly reflects reproductive isolation in pea aphid biotypes. Reproductive isolation determines effective gene flow in sympatric populations and hence their genetic differentiation at equilibrium between gene flow and drift (34). If equilibrium is not reached, genetic differentiation may mostly reflect divergence time, rather than effective gene flow. Likewise, its correlation with hybrid frequency could be explained by increasing reproductive isolation during divergence, as expected under the process of speciation. Variable spatial distance separating host species at all sampling sites may also account for the correlation between hybrid frequency and genetic divergence in pea aphid biotypes. However, the genetic structure of these biotypes appears moderately affected by distances of several hundreds of kilometers, because they constitute homogeneous genetic clusters at the scale of western Europe (Fig. 1*A*). Therefore, the distance separating host species at sampling sites, which was of hundreds of meters at most, would not explain the various levels of differentiation observed here (Fig. 2). The negative correlation between population differentiation and the proportion of hybrids among the 11 biotypes would thus portray various stages of speciation. Specifically, the 3 most differentiated biotypes A, B, and I, on *C. scoparius*, *Ononis* species, and *L. pratensis*, can be considered to be nearing complete speciation because no hybrid could be detected with any other sympatric biotype (Fig. 2). By contrast, the 8 other biotypes that presented various proportions of hybrids would constitute host-specialized races (11). They likely belong to the same species, given the diverse origins of

Table 1. Distribution of sampled pea aphids belonging to 11 biotypes on the surveyed plants

	Biotype												
Collection plant	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	Total	Migrants (%)
Cytisus scoparius	**156**											156	0.0
Ononis repens		**48**							1	1		50	4.0
O. spinosa		**77**	2									79	2.5
Lotus corniculatus			**56**	2	1		4	9		2	4	78	28.2
L. pedunculatus			5	**68**	1		2					76	10.5
Trifolium campestre					**22**		3					25	12.0
T. dubium					**12**							12	0.0
T. repens					**66**		4	2		4	1	77	14.3
T. pratense	1		1		**202**		3				3	210	3.8
Melilotus albus						**52**				3		55	5.5
M. officinalis						**54**		2			1	57	5.3
Pisum sativum							**139**					139	0.0
Vicia faba					2		**48**				4	54	11.1
V. sativa					1		**30**	5				36	16.7
V. hirsuta					6		**37**	1		2	4	50	26.0
V. cracca			2		1		7	**102**	1	1		114	10.5
Lathyrus pratensis						1		16	**97**	2		116	16.4
Medicago lupulina			1		1		2	1		**113**	4	122	7.4
M. sativa	2				1					5	**182**	190	4.2
Total	159	125	67	70	316	107	279	138	99	133	203	1696	9.4

Biotypes are denoted by the same letters as in Fig. 1. Numbers in boldface type indicate "resident" individuals, as opposed to "migrants" found on plants typical of other biotypes. To give a more accurate picture of migration patterns between plants, this distribution considers all aphids that could be clearly assigned to a biotype (thus excluding hybrids), counting individuals of the same parthenogenetic lineage (see *Methods*). Because all aphids were wingless and sampled on separate plants, these individuals represent different settlements.

their hybrids and lower pairwise genetic differentiation (Table S1). This hypothesis is supported by very low sequence polymorphism at noncoding nuclear makers, in comparison to interspecific divergence with a related species *A. kondoi* at the same loci (Fig. S2).

Implications for Ecological Divergence and Reproductive Isolation. Gene flow between populations often constrains their adaptive divergence (35). In the pea aphid, pairwise differences in host performance between biotypes do not correlate with their genetic differentiation (Fig. 3) and do not provide evidence that current hybridization hinders ecological divergence. Strong selection by host plants (ref. 19 and Fig. S1) may explain this observation, genetic differentiation being likely higher at loci controlling host adaptation compared to the average genetic divergence at our microsatellites (36). Also, western European races remain highly differentiated at many microsatellite markers (usually >50%, Fig. 2) despite appreciable hybridization. This observation strongly suggests that hybrids contribute only weakly to subsequent generations, in other words, postzygotic isolation. In North American host races, postzygotic isolation is the by-product of host plant adaptation, as trade-offs in host performance prevent the optimal use of parental plants by hybrid lineages (22). Postzygotic isolation may have an ecological basis in European biotypes, which present pronounced host specialization (ref. 19 and Fig. S1). They also produce hybrids

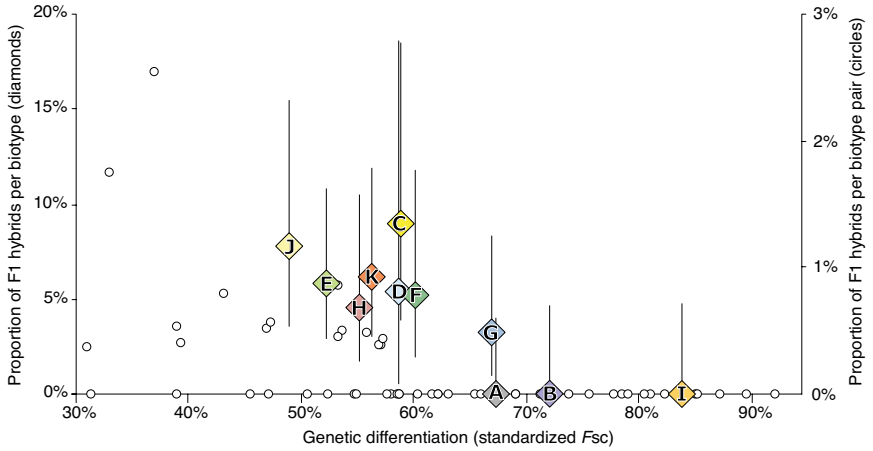

Fig. 2. Correlation of the proportion of hybrids with the genetic differentiation between biotypes of the pea aphid. The proportions of hybrids both per pair of biotypes (circles) and per biotype (diamonds, with 95% confidence intervals) correlate negatively with genetic differentiation (67.8% variance explained, $P < 0.001$, Mantel test). Letters denote biotypes as in Fig. 1. Genetic differentiation between 2 biotypes at microsatellite markers (standardized F_{SC}) was estimated within location by hierarchical analyses of molecular variance. For a given biotype (diamond), F_{SC} was averaged over the 10 pairwise comparisons involving this biotype.

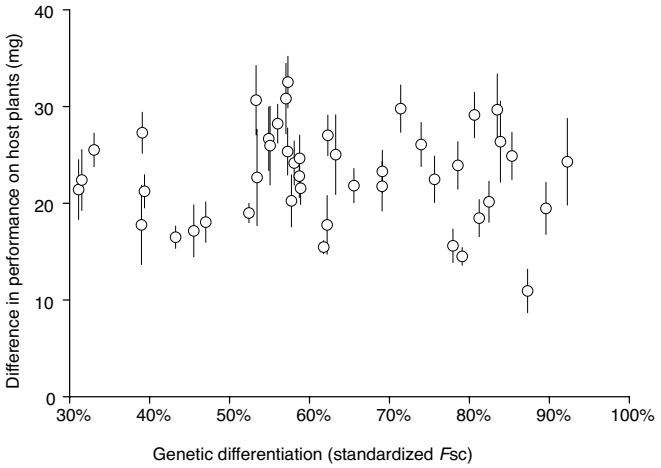

Fig. 3. Adaptive divergence between biotypes of the pea aphid does not correlate with their genetic differentiation. Adaptive divergence is based on host-specialization tests. It represents the Euclidian distance between 2 biotypes from their performance (produced biomass in milligrams) on their 2 representative host plants (see Fig. 1). Errors bars represent standard deviations over Euclidian distances that were computed between any 2 laboratory lineages of these biotypes.

underperforming on the parental plants, but growing normally on a favorable medium (37).

The detection of hybrids also demonstrates that moderate levels of migration between plant species (Table 1) can induce appreciable gene flow. Therefore, preference for different host species (17) and reduced performance on unfavorable plants (18) (Fig. S1), which limit migrations (38, 39), probably contribute to premating isolation in the 8 pea aphid host races. However, higher genetic differentiation in pea aphid biotypes is not related to higher difference in host performance (Fig. 3). Consequently, the apparent reduction of hybridization seen in the most differentiated biotypes (Fig. 2) may not reflect stronger host-mediated selection against migrants and hybrids. Late stages of speciation could instead reflect better host discrimination, possibly increasing habitat isolation on *C. scoparius* and *Ononis*, which harbored very few migrants (Table 1). It is also possible that nonecological premating barriers, such as behavioral isolation (40), are favored by avoiding maladaptive hybridization (41), enforcing the final course of speciation in sympatry.

Conclusion

Western European pea aphids exhibit 8 host races within a single species and belong to a larger complex of at least 11 sympatric host-specialized biotypes. Without knowledge on the past distribution of aphid populations and their host plants, the extent to which the divergence of pea aphids has been sympatric remains an unresolved question. However, the continuum of speciation formed by the 11 biotypes, specialized on host plants that are now broadly sympatric across Eurasia, strongly suggests that host races constitute a route for sympatric divergence in phytophagous insects. As yet, host races are documented only in a handful of insect species within the prodigious diversity of this class of animals. However, this apparent rarity may reflect mostly a lack of data on levels of hybridization in a number of good candidate models (11). Additionally, genetic studies on species that are thought to be homogenous may reveal unsuspected genetic and ecological diversity in sympatry (42).

Methods

Field Sampling. Host plants were chosen among common species known to harbor pea aphids (25). These plant species are widely distributed within western Europe and most of Eurasia (references in International Legume Database & Information Service; http://www.ildis.org/). They were sampled in grasslands and perturbed areas, where they usually grow in mixtures, or in adjacent monocultures of *Pisum*, *V. faba*, and some *M. sativa*. The main sampling location measured 40 km in diameter and was centered on Lantenay, eastern France (46°03′N, 5°32′E). This area was surveyed in August 2006 and in June and July 2007. Complementary sampling locations were the regions surrounding the towns of Le Rheu, northwestern France (48°06′N, 1°47′W) surveyed in June 2006 and Jena, located in eastern Germany (50°55′N, 11°35′E), surveyed in July 2006. In each location, wingless aphids from each plant species were sampled in at least 2 sites separated by a distance of several kilometers. No more than 1 individual per plant was sampled and then stored in 95% ethanol.

Genotyping and Sequencing. Aphids were genotyped with 2 multiplexes of 7 microsatellite loci (43–46) amplified and analyzed as in ref. 47. Table S2 details the loci used, the redesign of some primers, and specific conditions of PCR amplification.

In any sample set from the same host species and location, the probability that 2 individuals produced by sexual recombination had the same genotype at all loci, $P_{ID(SIB)}$ (49), computed with the program Gimlet (48), was $<0.1\%$. Given that aphids reproduce asexually from spring to autumn, this low probability indicated that identical 14-locus genotypes had resulted from parthenogenesis and not from sexual reproduction, which occurs during the fall. Including several individuals of the same genotype in the analyses would be equivalent to using the same individuals multiple times when analyzing strictly sexual populations. To circumvent this problem, we retained only 1 individual per 14-locus genotype (50) in the sample set where the corresponding genotype was most frequent.

Null alleles were suspected at 3 loci (Table S2) showing heterozygote deficit and a lack of amplification in rare individuals. These loci were discarded for analyses, except for assignment tests, which are only weakly affected by null alleles (51) and improve with the number of loci used (52).

We amplified and sequenced the noncoding, flanking regions of 3 microsatellite loci (Table S2) for each allele found in at least 1 homozygous individual (determined by the genotyping procedure) and in an outgroup, *A. kondoi*. We used a standard protocol described in ref. 47. This provided a comprehensive assessment of the genetic diversity found in the pea aphid at these loci, while direct sequencing avoided the risk of cloning PCR products bearing in vitro mutations. Sequences were aligned with the program BioEdit (53), coding contiguous missing nucleotides as single indels. Maximum-likelihood trees (Fig. S2) were built by heuristic searches in PAUP* 4.0b10 (54), using default settings and the most likely substitution model determined by Akaike information criteria in Modeltest 3.7 (55).

Detection of Host-Associated Populations. Using Structure 2.2 (26–28), we assigned individuals to inferred populations, assuming genetic admixture and correlated allele frequencies between them. We varied parameter k (the number of assumed populations) from 2 to 19. Twenty simulations were run for each value of k for 2×10^5 Markov Chain Monte Carlo (MCMC) steps, the first 10^5 being discarded. The posterior probability of the data at different values of k (56) did not outline a more likely number of populations (Fig. S3). We however retained the structuring solution of the run showing higher posterior probability at $k = 11$, because no other host-associated population was detected at higher values of k or when separately analyzing the structure of clusters B, E, F, and G, which are associated with several host species.

Detection of Hybrids. The parental origins of individuals were inferred one generation backward, with the ''use PopInfo'' option in Structure, which specifies a prior on their source population. Here, we used the collection plant of an individual as a prior and defined 11 putative populations by their host ranges, on the basis of results from the Structure analysis with admixture (Fig. 1*A*). We assumed prior migration rates of 10% between populations, as previously estimated between North American host races (39) (parameter ''MIGRPRIOR''). Other parameters were the same as specified in the admixture model at $k = 11$. Assignment tests on generated individuals of known parental origins presented low error rates ($<3\%$) under these settings, especially for highly differentiated biotypes (*SI Methods* and Fig. S4).

We then used the program NewHybrids (31) to identify hybrids in pairs of genetic clusters determined by Structure. We defined only 1 class of hybrids, excluding F_2 and backcrosses, and used Jeffrey's priors (31) as these settings proved the most accurate for simulated individuals. The accurate detection of F_2 and backcrosses usually requires many more loci (52). An individual was considered as a hybrid between 2 biotypes if it was assigned as such with a posterior probability $\geq 95\%$ in NewHybrids and had intermediate ancestry to

these biotypes in Structure. The parental origin of all individuals bearing different genotypes is shown in Table S1.

Assignment of Migrants. To genetically assign migrants with confidence, it was necessary to identify individuals potentially belonging to other undetected biotypes specialized on host plants that were not inspected. To detect these outliers, we computed the likelihoods of assignment of all individuals to 11 reference populations with Baudoin and Lebrun's Bayesian criterion (57) implemented in the program Geneclass 2 (30). On the basis of the output of Structure under the PopInfo model (see above), reference populations were defined as genetic clusters comprising only resident aphids (see Table 1), as these were less likely to be outliers. Migrants had likelihood values that were in the range of those of residents and thus showed no evidence that they belonged to undetected biotypes, with the exception of 3 individuals (Fig. S5). These presented clearly lower likelihoods of belonging to their reference population, and they were excluded from Table 1 and from further analyses. This analysis allowed us to delineate biotypes as groups of individuals assigned to a given host-associated population, i.e., to exclude outliers and F_1 hybrids.

Genetic Differentiation. Genetic differentiation (Figs. 2 and 3) was computed by hierarchical analyses of molecular variance (33) under Arlequin v. 3.11 (32). We structured the data set by location and then by biotype. Considering 2 biotypes at a single time, we computed the within-group (here, within-location) component of molecular variance, F_{SC}, which represents the mean genetic differentiation between these biotypes in sympatry. To correct for variations in genetic diversity at our markers, which may result from different mutation rates or effective population sizes among biotypes, F_{SC}'s were standardized according to Meirmans (58). F_{SC} values for all pairs of biotypes are shown in Table S1.

Host-Specialization Tests. Seventy individuals sampled from various plants were transferred to the laboratory and reared on separate broad bean plants (*V. faba*), a suitable host for all known biotypes of the pea aphid (18), to generate parthenogenetic lineages. None of them rejected this new host, but 3 of the aphids failed to develop lineages as they were parasitized by hymenoptera.

For performance tests, we selected a subsample of 45 aphid lineages assigned to 10 biotypes and bearing different genotypes, maximizing their diversity in terms of collection plant and location (Fig. S1). Lineages from the same host plant and location were chosen at random. No hybrid lineage was identified in our live collection for testing. Biotype C (on *L. corniculatus*) was not included because we initially expected the same biotype as that associated with *L. pedunculatus*. Hence, no plant of *L. corniculatus* was grown for performance tests.

Chosen lineages were maintained on broad bean plants at low density for at least 3 generations. Using a protocol modified slightly from ref. 47, we installed 10 first instar nymphs of each lineage on a potted plant (measuring at least 20 cm in height) in a climatic chamber (17 °C, 16-h photophase). After 9 days, surviving aphids, which had typically reached early adult stage, were weighed with a precision of 10^{-5} g. Three replications per lineage per plant species were carried out in the same climate chamber.

As a proxy for performance, we used the total biomass produced by a lineage *i* on a given plant species *a* (hereafter termed *Mia*), combining its survival and growth after 9 days, and averaged over the 3 replicates. By doing so, lineages were each described by 10 performance variables (on each test plant). A principal component analysis (PCA) was computed on this data set. The hierarchical ascendant classification (HAC) of lineages (Fig. 1*B*) was based on their coordinates on the 10 components of the PCA. This HAC used Ward's criterion of aggregation (59), which consists of merging iteratively the 2 lineages or groups of lineages with the lowest increase in intragroup variance and the lowest decrease in intergroup variance of the partition. This process yields a binary segmentation tree reflecting the hierarchy of similarities between lineages and wherein branch lengths represent the decrease of variance involved in the creation of a node. This analysis was done using SPAD 6 (http://spadsoft.com/).

The Euclidian distance between 2 lineages based on their performances when reared on their 2 typical host species was computed as

$$\sqrt{(Mia - Mja)^2 + (Mib - Mjb)^2}$$

whereby lineages *i* and *j* belong to different biotypes that are specialized on host plants *a* and *b*, respectively. The Euclidian distance between 2 biotypes (Fig. 3) was averaged over all pairwise comparisons between lineages.

The correlation between ecological and genetic differentiation (F_{SC}, previously computed in Arlequin) was tested using program FSTAT v. 2.9.3.2 (60).

ACKNOWLEDGMENTS. We thank Servane Penvern, Wolfgang Weisser, and Brigitte Pélisson for their contribution and logistical support in field sampling. Nancy Moran, James Mallet, Douglas Bailey, Thomas Baldwin, and 2 anonymous referees provided valuable comments on the manuscript. We also thank Lucie Mieuzet and Solène Coedel for their support in genotyping and Maurice Hullé and Guillaume Evanno for helping with the analysis of performance data. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation. This work was funded by program ECOGER, "Écologie pour la gestion des écosystèmes et de leurs ressources".

1. Schluter D (2001) Ecology and the origin of species. *Trends Ecol Evol* 16(7):372–380.
2. Coyne JA, Orr HA (2004) *Speciation* (Sinauer Associates, Sunderland, MA), p 545.
3. McKinnon JS, et al. (2004) Evidence for ecology's role in speciation. *Nature* 429(6989):294–298.
4. Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc Natl Acad Sci USA* 103(9):3209–3213.
5. Rundle HD, Nosil P (2005) Ecological speciation. *Ecol Lett* 8(3):336–352.
6. Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* 323(5915):737–741.
7. Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol* 16(7):381–390.
8. Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Evol Syst* 38(1):459–487.
9. Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos Trans R Soc Lond Ser B Biol Sci* 363(1506):2971–2986.
10. Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annu Rev Entomol* 47:773–815.
11. Drès M, Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philos Trans R Soc B Biol Sci* 357(1420):471–492.
12. Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). *Evolution* 23:237–251.
13. Phillips PA, Barnes MM (1975) Host race formation among sympatric apple, walnut and plum populations of *Cydia pomonella*. *Ann Entomol Soc Am* 68:1053–1060.
14. Carroll SP, Boyd C (1992) Host race radiation in the soapberry bug: natural history with the history. *Evolution* 46(4):1052–1069.
15. Müller FP, Steiner H (1985) The problem *Acyrthosiphon pisum* (Homoptera: Aphididae) (translated from German). *Z Angew Zool* 72:371–334 (in German).
16. Caillaud MC, Via S (2000) Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *Am Nat* 156(6):606–621.
17. Ferrari J, Godfray HCJ, Faulconbridge AS, Prior K, Via S (2006) Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution* 60(8):1574–1584.
18. Ferrari J, Via S, Godfray HCJ (2008) Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution* 62(10):2508–2524.
19. Via S (1991) The genetic structure of host plant adaptation in a spatial patchwork: Demographic variability among reciprocally transplanted pea aphid clones. *Evolution* 45:827–852.
20. Müller FP (1964) Biotypes and subspecies of the ''pea aphid'' *Acyrthosiphon pisum* (Harris) (translated from German). *Z Pflanzenkr* 69:129–136 (in German).
21. Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* 53(5):1446–1457.
22. Hawthorne DJ, Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412(6850):904–907.
23. Simon JC, et al. (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proc Biol Sci* 270(1525):1703–1712.
24. Frantz A, Plantegenest M, Mieuzet L, Simon JC (2006) Ecological specialization correlates with genotypic differentiation in sympatric host-populations of the pea aphid. *J Evol Biol* 19(2):392–401.
25. Eastop VF (1971) Keys for the identification of *Acyrthosiphon* (Hemiptera : Aphididae). *Bull Br Mus Nat Hist Entomol* 26(1):1–115.
26. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
27. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7(4):574–578.
28. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.
29. Dixon AFG (1997) *Aphid Ecology—An Optimization Approach* (Kluwer, Norwell, MA), 2nd Ed.
30. Piry S, et al. (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *J Hered* 95(6):536–539.
31. Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160(3):1217–1229.

32. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
33. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes—application to human mitochondrial DNA restriction data. *Genetics* 131(2):479–491.
34. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16(2):97–159.
35. Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecol Lett* 11(6):624–636.
36. Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol* 17(19):4334–4345.
37. Müller FP (1971) Isolating mechanisms between sympatric races of the pea aphid *Acyrthosiphon pisum* (Harris) (Homoptera, Aphididae) (translated from German). *Zool Jb Syst* 98:131–152 (in German).
38. Nosil P, Vines TH, Funk DJ (2005) Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* 59(4):705–719.
39. Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* 54(5):1626–1637.
40. Margaritopoulos JT, Malarky G, Tsitsipis JA, Blackman RL (2007) Microsatellite DNA and behavioural studies provide evidence of host-mediated speciation in *Myzus persicae* (Hemiptera : Aphididae). *Biol J Linn Soc* 91(4):687–702.
41. Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: theory and data. *Annu Rev Ecol Evol Syst* 34(1):339–364.
42. Stireman JO, Nason JD, Heard SB (2005) Host-associated genetic differentiation in phytophagous insects: General phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution* 59(12):2573–2587.
43. Caillaud MC, et al. (2004) Microsatellite DNA markers for the pea aphid *Acyrthosiphon pisum. Mol Ecol Notes* 4(3):446–448.
44. Wilson ACC, et al. (2004) Cross-species amplification of microsatellite loci in aphids: assessment and application. *Mol Ecol Notes* 4(1):104–109.
45. Simon JC, et al. (1999) Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Mol Ecol* 8(4):531–545.
46. Kurokawa T, Yao I, Akimoto SI, Hasegawa E (2004) Isolation of six microsatellite markers from the pea aphid, *Acyrthosiphon pisum* (Homoptera, Aphididae). *Mol Ecol Notes* 4(3):523–524.
47. Peccoud J, et al. (2008) Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. *Mol Ecol* 17(21):4608–4618.
48. Valiere N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Mol Ecol Notes* 2(3):377–379.
49. Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. *Mol Ecol* 10(1):249–256.
50. Halkett F, Simon JC, Balloux F (2005) Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol Evol* 20(4):194–201.
51. Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. *J Hered* 99:616–623.
52. Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Mol Ecol* 15(1):63–72.
53. Hall T (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98.
54. Swofford DL (2000) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods*) (Sinauer Associates, Sunderland, MA).
55. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14(9):817–818.
56. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14(8):2611–2620.
57. Baudouin L, Lebrun P (2001) An operational Bayesian approach for the identification of sexually reproduced cross-fertilized populations using molecular markers. *Acta Hort* 546:81–93.
58. Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60(11):2399–2402.
59. Ward JJH (1963) Hierarchical grouping to optimise an objective function. *J Am Stat Assoc* 58(301):236–244.
60. Goudet J (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. *J Hered* 86(6):485–486.