Supporting Information

Peccoud et al. 10.1073/pnas.0811117106

SI text

To assess the reliability of Structure to correctly identify hybrids, we assigned the simulated genotypes of individuals of known parental origins. First-generation (F_1) hybrids were generated by random allele draws from the allele frequency distributions of 2 biotypes, which were provided by the output of Structure at $k = 11$. "Purebreds" (F_0) were generated by draws from the allelic frequencies of a single biotype. For assignments using the "PopInfo" model, F_1's were randomly considered to come from either parental population (here, host plant). We simulated moderate hybridization between biotypes by adding 5 F_1's of each origin (275 genotypes for 55 types of cross) to the microsatellite data set of 1,090 real individuals. We also added 5 F_0's of each of the 11 biotypes separately, constituting another data set. The inclusion of 5 genotypes per parental origin allowed us to assign a sufficient number of generated genotypes per run, which did not visibly alter the ancestry values of real individuals.

In these data sets, assignment tests indicated better performance of the PopInfo method of Structure, compared to the admixture model [supporting information (*SI*) Fig. S4*A*].

Forty similar data sets with different simulated genotypes (100 for each parental origin) were then created to better estimate the risk of assignment errors by the PopInfo method. The most common error, which occurred in less than 3% of the generated genotypes, was the assignment of an F_1 to a single one of its parental biotypes (i.e., with an ancestry value of $\geq$90%) and thus a possible attribution as an F_0 (Fig. S4*B*). By contrast, almost all F_0's were assigned to a single cluster by Structure. We concluded that the assignment method was unlikely to overestimate hybridization.

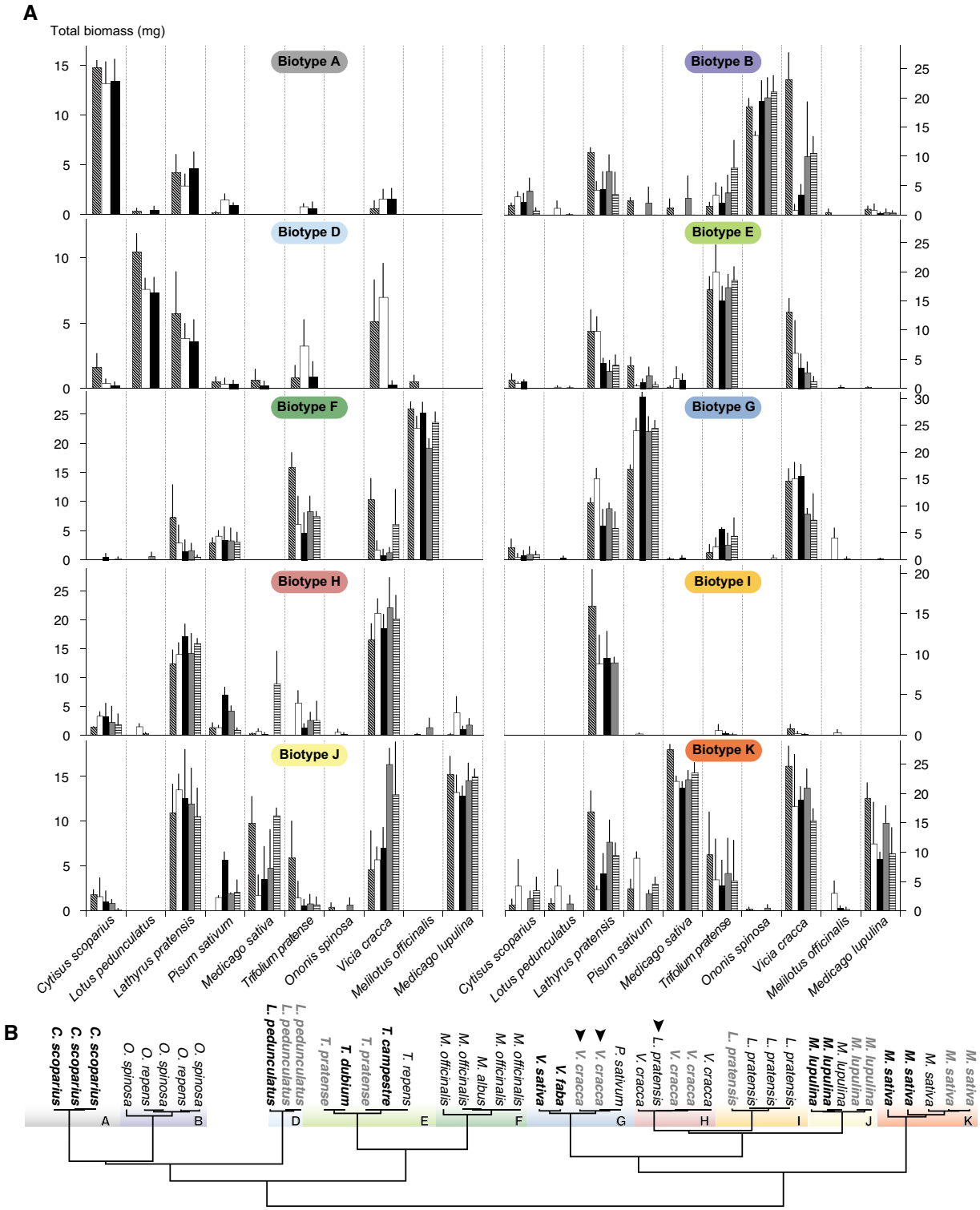


Fig. S1. (*A*) Performance (total biomass of surviving aphids, averaged over replicates) of 45 lineages from 10 biotypes of the pea aphid, when reared on 10 plant species (Latin names, *x* axis). Bars of different shading correspond to different lineages within a given biotype; errors bars represent standard errors over replicates. (*B*) Phenogram (see Fig. 1*B*, main text) with indications of collection plants and locations of test lineages. Regular type, Lantenay; boldface type, Le Rheu; shaded type, Jena. For a given biotype, lineages appear in the same order, from left to right, in the phenogram and the histogram. Arrowheads indicate lineages that were initiated by migrant aphids, i.e., individuals genetically assigned to biotypes that are not associated with their collection plants.

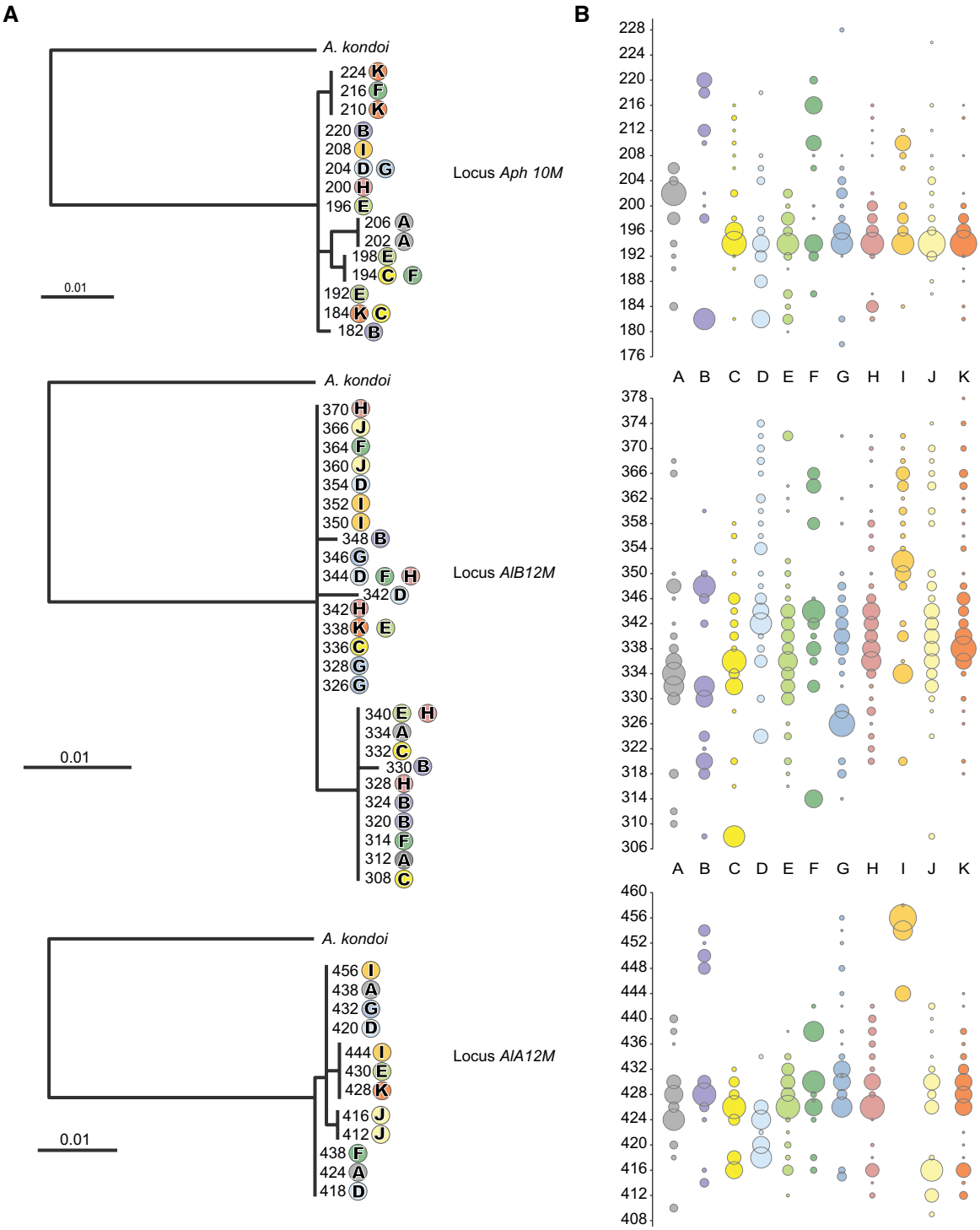


Fig. S2. (*A*) Maximum-likelihood trees based on sequence variation in 3 microsatellite flanking regions (Table S2) among biotypes of the pea aphid and an outgroup. Numbers refer to alleles (sizes in base pairs estimated by the genotyping procedure). Colors and letters represent biotypes, as in the main text. These noncoding sequences represent short genetic divergence between biotypes, compared to the divergence with *A. kondoi*, which suggests their recent diversification. (*B*) Distributions of microsatellite alleles in the 11 biotypes, for the corresponding loci. Circled areas are proportional to allele frequencies.

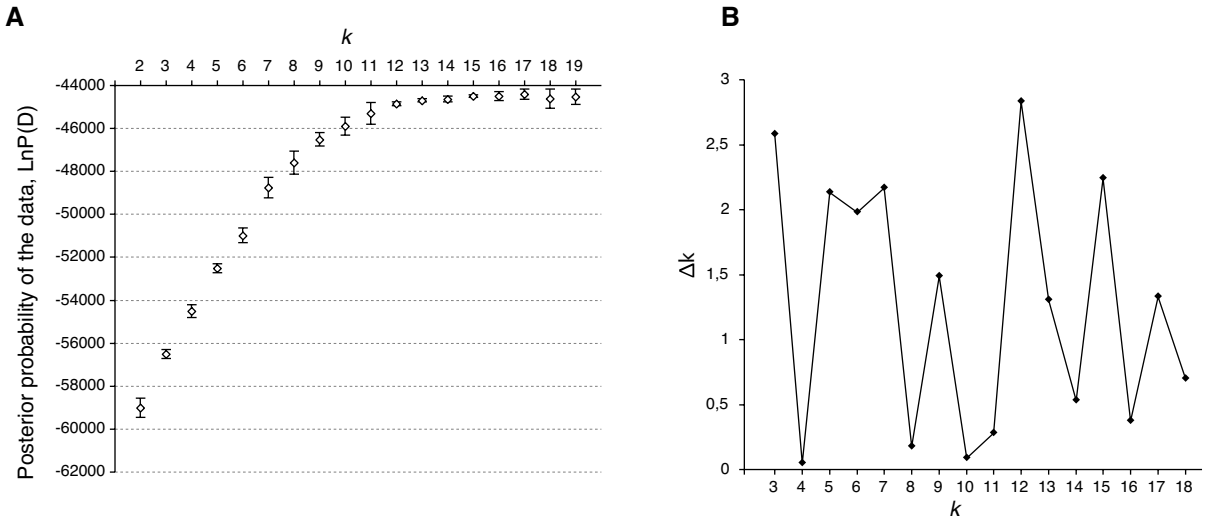
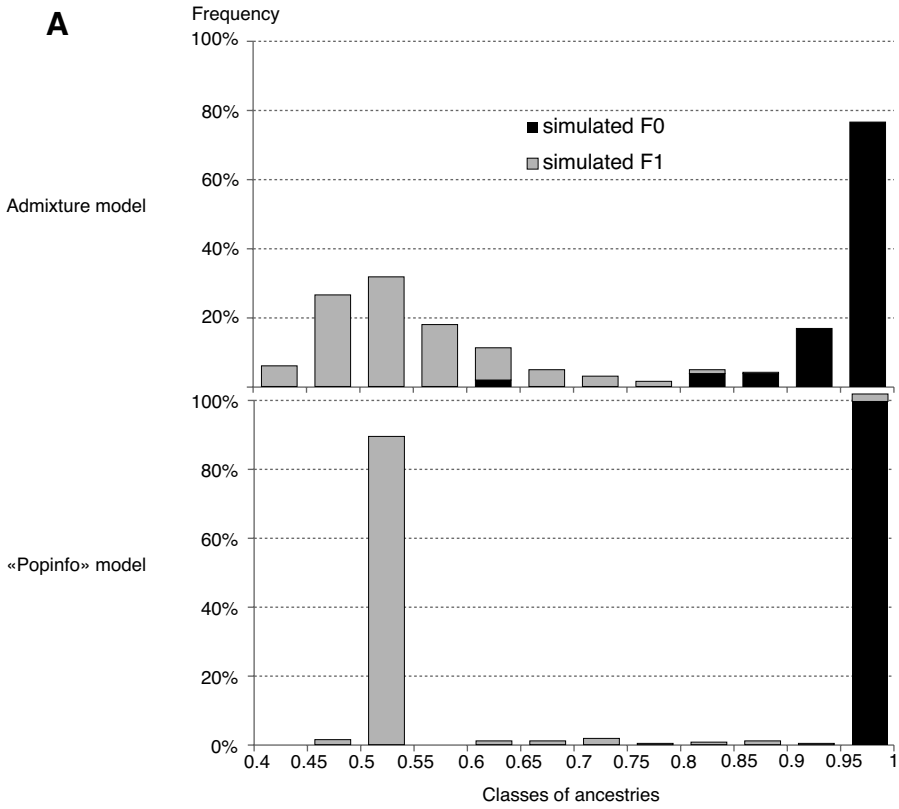


Fig. S3. (*A*) Logarithm of the posterior probability of the data, Ln *P*(*D*), given the number of assumed populations in the pea aphid (parameter *k*), as estimated by Structure. Probabilities are averaged over 20 runs per value of *k*, and error bars represent the standard deviation in Ln *P*(*D*) over these runs. (*B*) The second-order rate of change of Ln *P*(*D*) in respect to *k* (Δ*k*, see ref. 57) does not outline a more likely number of populations, for which Δ*k* would have been clearly higher.

A



B

	Parent biotype											
Observations	A	B	C	D	E	F	G	H	I	J	K	averaged
F1 of ancestry > 0.9	-	-	0.7%	1.4%	6.8%	0.8%	1.3%	5.7%	-	5.6%	8.9%	2.84%
F0 of ancestry < 0.9	-	-	-	-	-	-	-	1%	-	-	-	0.09%
Assignement error	0.10%	-	0.4%	0.4%	1.6%	0.5%	0.1%	3.1%	-	1.7%	3.7%	1.29%

Fig. S4. (*A*) Comparison of the ability of 2 clustering models implemented in Structure to discriminate between generated F_0 and F_1 crosses among 11 biotypes of the pea aphid: the admixture model and the ''PopInfo'' model (see *Methods*). Histograms show the distributions of individual ancestries (*q*-values) for simulated genotypes, considering only the highest *q*-value for each genotype. An F_1 is expected to present ancestries of 50% to each of its parental populations, and an F_0 should be assigned at 100% to its parental population. Results obtained with the PopInfo model are closer to these expectations than those obtained with the admixture model. (*B*) Assignment errors of Structure on 1,100 simulated F_0 and 5,500 F_1 genotypes, using the PopInfo model. The first kind of observation corresponds to F_1's assigned by the program mostly to a single biotype. The second kind of observation corresponds to F_0's that were assigned with low ancestry to their biotype. These observations may respectively cause an underestimation and an overestimation of hybridization. The last kind corresponds to F_1's having their highest or second highest ancestry value assigned to neither of their true parent biotypes. These errors did not affect F_0 genotypes, and they were scored in the parental biotypes to which individuals should have been assigned.

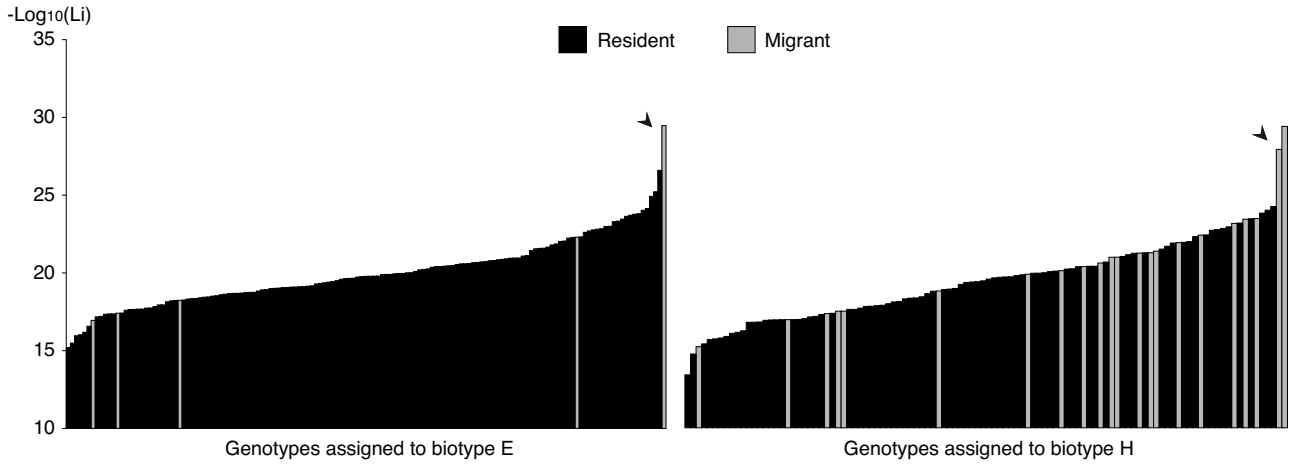


Fig. S5. Genotypes ranked according to the $-\mathrm{Log}_{10}$ of the likelihood that they belonged to their reference population (biotype), as computed by the program Geneclass (see *Methods*). Hybrids are not included. ''Residents'' correspond to individuals that were assigned to the reference population that is associated with their collection plant, as opposed to ''migrants'' (see Table 1, main text). Arrowheads indicate ''outliers,'' which showed higher $-\mathrm{Log}_{10}$(Li) than all residents. These outliers were statistically assigned to biotypes E and H, but they may actually belong to other undetected biotypes.

Table S1. Individual ancestries and genetic differentiation in 11 biotypes of the pea aphid

	Biotype										
	J	**E**	**H**	**K**	**D**	**C**	**F**	**G**	A	B	I
J	**84**										
E	**1**	**146**									
H		**1**	**105**								
K		**5**		**122**							
D	**1**	**1**			**35**						
C	**4**		**1**	**1**		**61**					
F	**1**	**1**	**1**	**1**			**92**				
G			**2**	**1**			**1**	**120**			
A									110		
B										95	
I											93
J	—	**39.4**	**45.5**	**39.0**	**43.3**	**37.2**	**39.1**	**55.1**	52.4	58.7	79.1
E	**10.7**	—	**31.1**	**33.1**	**47.0**	**50.7**	**57.3**	**54.9**	58.9	69.1	81.2
H	**12.4**	**7.1**	—	**31.5**	**62.2**	**47.5**	**56.0**	**53.4**	65.6	75.6	82.4
K	**11.1**	**8.1**	**7.6**	—	**58.1**	**53.8**	**57.3**	**57.1**	74.0	80.6	78.6
D	**14.9**	**13.7**	**18.0**	**17.5**	—	**47.2**	**58.8**	**63.3**	61.8	57.7	87.3
C	**12.9**	**14.8**	**14.3**	**16.6**	**17.9**	—	**60.5**	**72.5**	67.2	66.2	85.1
F	**13.5**	**16.7**	**16.9**	**17.7**	**22.0**	**23.0**	—	**53.3**	62.3	71.4	85.3
G	**18.2**	**15.4**	**14.9**	**16.5**	**21.4**	**26.2**	**19.6**	—	83.9	83.5	92.2
A	21.0	20.6	23.0	26.6	27.0	29.2	26.4	34.5	—	69.1	77.9
B	23.5	23.5	26.7	29.0	25.5	29.1	30.6	34.8	32.8	—	89.6
I	28.0	24.2	25.5	24.8	33.5	33.1	32.7	34.6	33.6	39.0	—

Upper half: distribution of pea aphids (one per microsatellite genotype, see *Methods*) according to their parental origin, one generation backward, as inferred by Structure and verified by NewHybrids (see *Methods*). Individuals occupy the diagonal, except F_1 hybrids, which are indicated below the diagonal only. Lower half: pairwise genetic differentiation (in percent) between biotypes, computed by hierarchical analyses of molecular variance. Above diagonal: standardized F_{SC}. Below diagonal: raw F_{SC}. All F_{SC}'s are significant (20,000 bootstraps over loci). Values are shown in boldface type for host races likely belonging to the same species.

Table S2. Dinucleotide microsatellite loci that were used in this study

Locus (reference)	Primer sequences (5′–3′) (see references in first column if not indicated)	Concentration used (nM) (see ref. 5 if not indicated)
*AlA09M** (1)		
AlB07M (1)		
AlB08M (1)		
AlB12M (1)	AAAACCCGTTGAAAATGGTG (F)	60
	(R)	60
*ApF08M** (1)		
ApH 08M (1)		
ApH 10M (1)		
AlA12M (1)	(F)	92
	(R)	92
AlB04M (1)	GGACTGAGGAACTCGAAACG (F)	60
	(R)	60
Ap-03 (2)	GCAGCAACAGCAGGTGTAAA (F)	60
	(R)	60
S23 (3)	(F)	92
	(R)	92
S30 (3)	CGATCCGACACAAAACACAC (F)	60
	CGTTTCGACTCTGCGTTGT (R)	60
*S3.43** (3)	(F)	92
	(R)	92
Sm11 (4)	GGTGATGGTGGCGTGAAC (F)	60
	ACAGACGGTGTCCGTAGTCC (F)	60
AlA12M (1)	TGTCTGATGCGCTTACGTTT (F)	
AlB12M (1)	CGGGTGCAGGGTATAAGGTA (F)	
ApH 10M (1)	TTGCTGACGACTTCAACTGC (R)	

The first 14 loci correspond to the multiplexes that were used for genotyping and the last 3 rows corresponds to the flanking regions that were sequenced (*SI* Fig. S2). The second multiplex (loci *AlA12M* to *Sm11*) was used as in ref. 5, with an annealing temperature of 60°C. F, forward primer; R, reverse primer. *, locus suspected to have null alleles (see *Methods*).

1. Caillaud MC, et al. (2004) Microsatellite DNA markers for the pea aphid *Acyrthosiphon pisum*. *Mol Ecol Notes* 4(3):446–448.
2. 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.
3. Wilson ACC, et al. (2004) Cross-species amplification of microsatellite loci in aphids: assessment and application. *Mol Ecol Notes* 4(1):104–109.
4. 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.
5. 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.