doi:10.1111/j.1558-5646.2012.01612.x

LARGE-SCALE CANDIDATE GENE SCAN REVEALS THE ROLE OF CHEMORECEPTOR GENES IN HOST PLANT SPECIALIZATION AND SPECIATION IN THE PEA APHID

Carole M. Smadja,[1,2] Björn Canbäck,[3] Renaud Vitalis,[4] Mathieu Gautier,[5] Julia Ferrari,[6] Jing-Jiang Zhou,[7] and Roger K. Butlin[8]

[1]Centre National de la Recherche Scientifique CNRS—Institut des Sciences de l'Evolution UMR 5554, cc065 Université Montpellier 2, 34095 Montpellier, France

[2]E-mail: carole.smadia@univ-montp2.fr

[3]Microbial Ecology, Department of Biology, Lund University, SE-223 62 Lund, Sweden

[4]Centre National de la Recherche Scientifique—Institut National de la Recherche Agronomique, UMR CBGP (INRA—IRD—CIRAD—Montpellier SupAgro), Campus International de Baillarguet, CS 30016, F-34988 Montferrier sur Lez Cedex, France

[5]Institut National de la Recherche Agronomique, UMR CBGP (INRA—IRD—CIRAD—Montpellier SupAgro), Campus International de Baillarguet, CS 30016, F-34988 Montferrier sur Lez Cedex, France

[6]Department of Biology, University of York, York YO10 5DD, United Kingdom

[7]Department of Biological Chemistry, Rothamsted Research, BBSCR, Harpenden, Hertfordshire AL5 2JQ, United Kingdom

[8]Animal and Plant Sciences, University of Sheffield S10 2TN, United Kingdom

Received December 16, 2011
Accepted February 13, 2012

Understanding the drivers of speciation is critical to interpreting patterns of biodiversity. The identification of the genetic changes underlying adaptation and reproductive isolation is necessary to link barriers to gene flow to the causal origins of divergence. Here, we present a novel approach to the genetics of speciation, which should complement the commonly used approaches of quantitative trait locus mapping and genome-wide scans for selection. We present a large-scale candidate gene approach by means of sequence capture, applied to identifying the genetic changes underlying reproductive isolation in the pea aphid, a model system for the study of ecological speciation. Targeted resequencing enabled us to scale up the candidate gene approach, specifically testing for the role of chemosensory gene families in host plant specialization. Screening for the signature of divergence under selection at 172 candidate and noncandidate loci, we revealed a handful of loci that show high levels of differentiation among host races, which almost all correspond to odorant and gustatory receptor genes. This study offers the first indication that some chemoreceptor genes, often tightly linked together in the genome, could play a key role in local adaptation and reproductive isolation in the pea aphid and potentially other phytophagous insects. Our approach opens a new route toward the functional genomics of ecological speciation.

KEY WORDS: **Gene flow, genome scan, genomic islands, natural selection, odorant receptor, targeted resequencing.**

EVOLUTION INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

Understanding the drivers of speciation is critical to interpreting patterns of biological diversity (Butlin et al. 2009). It now seems likely that natural selection plays a key role in the speciation process in many, perhaps most cases (Kirkpatrick and Ravigné 2002) and that selection is capable of promoting the evolution of reproductive isolation in the face of gene flow (Servedio and Noor 2003; Gavrilets 2004; Bolnick and Fitzpatrick 2007; Smadja and Butlin 2011). The emphasis of speciation research has thus shifted away from debates about geographic modes (sympatric/parapatric/allopatric speciation) (Butlin et al. 2008) and toward the dissection of the ecological and genetic mechanisms of speciation (e.g., Elias et al. 2008; Lowry and Willis 2010).

Since the influential paper by Wu (2001), there has been a realization that, when reproductive isolation is not complete, patterns of differentiation are likely to be uneven across the genome (Nosil et al. 2009). In the case of ecologically driven speciation with gene flow, reproductive isolation is likely to start with divergent selection on loci that contribute to local adaptation. Once divergence has begun, the completion of speciation requires the initial isolation to be augmented by further adaptation, habitat choice or assortative mating, the evolution of premating isolation being most likely where there is either pleiotropy or close physical linkage between the locally adapted alleles and alleles that increase nonrandom mating (Felsenstein 1981; Servedio 2009; Smadja and Butlin 2011). Therefore, we expect isolation to spread, genomically, from its initial foci to larger genomic regions and up to the whole genome, as the degree of reproductive isolation increases. Yet, the details of the genetic architecture and dynamics of reproductive isolation remain poorly understood and documented (Nosil and Schluter 2011).

So far, two main approaches have been developed to address the genetics of speciation. On the one hand, population genomic approaches, which use allele frequency comparisons among taxa to detect "outlier" loci showing some sign of divergence under selection, have typically provided evidence for a few percent of the genomic regions covered by markers to be maintained at higher levels of differentiation than the background balance between gene flow and genetic drift (reviewed by Butlin et al. 2008; Nosil et al. 2009 among others). On the other hand, quantitative trait locus (QTL) analysis, which uses phenotypic and genotypic data to identify loci explaining variation in traits, has been widely used to address the genetic basis of reproductive isolation (e.g., Baxter et al. 2009; Shaw and Lesnick 2009). This strategy can be particularly powerful for identifying regions of the genome involved in reproductive isolation when combined with the population genomics approach (Stinchcombe and Hoekstra 2008), a combination that has been applied to whitefish (Rogers and Bernatchez 2007) and pea aphids (Via and West 2008). However, although population genomics and QTL approaches help in gaining insights into the number, size, and distribution of differentiated genomic regions between partially isolated populations, the rather large and/or anonymous regions they point at often make more precise identification of the key loci and genetic changes involved in reproductive isolation impractical solely by further application of these methods.

A CANDIDATE GENE APPROACH USING CHEMOSENSORY GENES AS A MODEL

In this article, we present a candidate gene approach, applied here to the pea aphid system, as an alternative route to the genetics of speciation, and which should complement the QTL and population genomic methods. The candidate gene approach has been used with great success in other contexts (e.g., in disease genetics, pharmacogenomics, animal, and plant breeding—reviewed in Zhu and Zhao 2007) but has rarely been applied in speciation research. When it has been applied, it has typically been in respect of a small number of genes for specific phenotypic traits (Lexer et al. 2004; Kronforst et al. 2006; Haas et al. 2009). In contrast, we here propose a large-scale candidate gene approach that focuses on gene families, allowing many loci, whose function can be inferred at least in general terms, to be compared with background differentiation in a set of randomly selected loci. Here, we focus on gene families with great potential for influencing speciation in many systems: the chemosensory genes (Dulac and Torello 2003; Hallem et al. 2006; Nei et al. 2008; Touhara and Vosshall 2009; Croset et al. 2010). The chemical senses are frequently involved in at least some aspects of premating isolation, especially host/habitat and mate choice (Smadja and Butlin 2009). This implies that chemosensory genes, and in particular gene products involved in peripheral processes of semiochemical recognition, are good candidates for contributing to premating isolation in many systems. Moreover, the genomic organization of these genes in very large multigene families (Kent and Robertson 2009; Niimura 2009; Sanchez-Gracia et al. 2009; Robertson et al. 2010), their mode of evolution under a birth-and-death model and evidence for positive selection in some branches of these multigene families, have led us and others to hypothesize a role for these chemosensory genes in host adaptation and specialization in insects (Matsuo et al. 2007; McBride 2007; McBride et al. 2007; Gardiner et al. 2008; Matsuo 2008; Dworkin and Jones 2009; Smadja et al. 2009; Schymura et al. 2010). Chemosensory genes could play a major role in aspects of both local adaptation and premating isolation. If some of these candidate genes are effectively involved in reproductive isolation, they should show high levels of differentiation among partially isolated taxa. This provides a clear and testable prediction that we here use to address the genetic basis of reproductive isolation and local adaptation in the pea aphid.

THE PEA APHID

The pea aphid, *Acyrthosiphon pisum*, is a well-established model system for ecological specialization and speciation (Via 2009; Peccoud and Simon 2010). Indeed, the pea aphid has been shown to form host-associated populations on different legume species (Via 1991; Ferrari et al. 2006; Ferrari et al. 2008; Peccoud et al. 2009), which have higher performance on and preference for the plant species that they have been found on in the wild compared to alternative host plants (Via 1991, 1999; Ferrari et al. 2006; Ferrari et al. 2008). Host-associated populations are genetically differentiated (Via 1999; Frantz et al. 2006; Peccoud et al. 2009) and host plant specialization is the key component of reduced gene flow in this system as it induces selection against immigrants and against hybrids (Via et al. 2000), as well as assortative mating, because pea aphids have no host alternation and reproduce on their preferred plant (Via 1999).

A QTL mapping analysis of host acceptance and performance in North American populations (Hawthorne and Via 2001) suggests a polygenic basis for those traits and a recent scan for selection based on amplified fragment length polymorphism markers showed that these QTL for major traits underlying reproductive isolation correspond to differentiated outliers (Via and West 2008). Hawthorne and Via's (2001) QTL analysis also suggested close physical linkage between loci controlling the two traits and/or pleiotropy. This suggests that host acceptance evolves under direct selection in pea aphids, a very favorable scenario for speciation with gene flow (Smadja and Butlin 2011). Moreover, host detection and feeding behaviors of pea aphids are closely coupled, as they do not identify their home plant from a distance. Indeed, host acceptance (or rejection) is essentially a question of taste and smell in aphids, the key step being the probing of epidermal plant tissues (Caillaud and Via 2000).

This information on the physiological basis of host acceptance in aphids led us to choose genes involved in chemoreception as primary candidates for host plant specialization and speciation in this system. Critically, the annotation work we completed on odorant (OR) and gustatory receptors (GR) (Smadja et al. 2009) and odorant binding and related proteins (OBP and CSP) (Zhou et al. 2010) in the recently sequenced pea aphid genome (Consortium, T. I. A. G 2010) provided an exceptional opportunity to investigate the genetic basis of host race formation using a candidate gene approach.

The specific objective of this study was to test for a role of chemosensory genes in the formation of three host races of pea aphids, which are closely related but highly specialized on different host plants from *Lotus pedunculatus*, *Medicago sativa*, and *Trifolium pratense* (Ferrari et al. 2008; Peccoud et al. 2009; Ferrari et al. 2011). Our hypothesis was that divergent selection must have operated on a small subset of chemosensory genes during the formation of these host races. We predicted that this selection would have generated patterns of sequence divergence at those loci that are atypical with respect to background divergence in the majority of chemosensory loci and in randomly selected loci with no known chemoreception function.

To solve the challenge of assessing sequence variation at numerous candidate loci (∼200) and in several populations, we used sequence capture technology coupled with next-generation sequencing (NGS) to analyze targeted sequence information and to scale up the candidate gene analysis. To our knowledge, this approach has not previously been used in the context of speciation research but has enormous potential.

Our approach involved the following steps:

(1) Sample three different host races from a small geographical area.
(2) Capture chemosensory genes and nonchemosensory genes (randomly chosen genes) from genomic DNA using NimbleGen capture arrays and sequence using the GS-FLX Titanium platform.
(3) Analyze these sequences for patterns of diversity, divergence, and signatures of selection.

The data obtained allow an unprecedentedly detailed picture of the genomic divergence between host plant races based on sequence data for a sample of ∼200 target loci (Table S1). Narrowing down the number of genes suspected to underlie adaptation and reproductive isolation in the pea aphid to a handful of loci of known functional class, this large-scale candidate gene analysis provides key information on the identity of loci and genetic changes involved.

Material and Methods

SAMPLES

Aphid sampling and rearing

We sampled pea aphids from three different host plants: greater birdsfoot trefoil *L. pedunculatus*, alfalfa *M. sativa*, and red clover *T. pratense*. These samples represent three different host races (Ferrari et al. 2008; Peccoud et al. 2009; Ferrari et al. 2011). Homogeneity in host plant acceptance and performance across large geographical areas within each host race and low spatial genetic structure suggest that samples restricted to a particular area are representative of the races across a larger geographic scale (Ferrari et al. 2011). Insects were collected from an area of 25-km radius centered at Silwood Park, Berkshire, UK (51°9¢30²N; 0°38¢15²W), and brought back to the laboratory. Each genotype was used previously and tested for performance on eight different plant genera to verify its specialization on the plant it was collected from (Ferrari et al. 2008). Clonal cultures were established from single individuals collected in the field and maintained under summer conditions in which only asexual offspring are produced

(maintenance on Vicia faba ("The Sutton" cultivar) at 15°C, 70% r.h., and a 16:8 h light:dark cycle within the same controlled environment room, Ferrari et al. 2006; Ferrari et al. 2008). Individual adult aphids from each genotype were regularly collected and stored in alcohol for future DNA extraction.

DNA sample preparation

In each population we extracted genomic DNA from eight aphid genotypes (i.e., we sampled 16 chromosome sets from each population) using DNeasy tissue kit (Qiagen, Crawley, West Sussex, UK) according to the manufacturer's protocol. As aphids produce apomictic clones, we used two individual aphids per clonal genotype to get enough DNA material. Extractions from each genotype were quantified fluorometrically using SYBR green on 1% agarose gels and then combined in equimolar mixture to constitute three population samples (pools). For each population sample, a minimum of 21 μg of unamplified high-quality genomic DNA was obtained.

SEQUENCE CAPTURE, SEQUENCING, AND SEQUENCE ANALYSIS

Exon array design

We used NimbleGen capture array technology (Roche) to perform targeted exon enrichment (Hodges et al. 2007). Primary sequence data for all target genes were extracted from Assembly 1.0 of the pea aphid genome using Apollo-AphidBase (Consortium 2010) and from detailed annotation information on the pea aphid chemosensory genes chosen here as candidate genes (Smadja et al. 2009; Zhou et al. 2010). We selected all of the chemosensory genes (odorant receptors [*Or*], gustatory receptors [*Gr*], odorant binding proteins [*Obp*], and ChemoSensory proteins [*Csp*]) that had been partially or fully annotated in Assembly 1.0 of the pea aphid genome (Smadja et al. 2009; Zhou et al. 2010). We also included noncandidate genes as targets, randomly chosen in the genome and a priori not involved in chemosensory processes. The capture target represented 1021 exons from 203 genes, including both candidate and noncandidate loci (see Tables S1 and S2A). NimbleGen bioinformatics service then completed the array design using default settings (Hodges et al. 2007) except for uniqueness, which was relaxed so that closely related paralogs could be captured. Overlapping microarray probes were designed to span each target region. We checked the array design using SignalMap version 1.9 (© 2006 NimbleGen System, Inc.).

Capture and sequencing

All of the capture and sequencing steps were performed by Lisa Olohan at NBAF-Liverpool (Natural Environment Research Council [NERC] Biomolecular Analysis Facility), United Kingdom. Quality control was applied to all three DNA samples, hereafter named *Lotus*, *Medicago*, and *Trifolium* in reference to the aphid host plants. For each sample, library preparation, capture on arrays and amplification was performed following the manufacturer's protocol. The success of enrichment was measured by qPCR at control loci. Captured fragments from each population were then sequenced using a Genome Sequencer FLX Titanium Instrument from Roche Applied Science (454 sequencing) according to the manufacturer's instructions. We used 1/8 of a 454 plate for each population, with repeats as necessary to give at least 75,000 reads per sample. All sequences were submitted to the short read archive [SRA048197.1].

Methods used for raw sequence analysis (read mapping, alignment quality control, single nucleotide polymorphism (SNP) calling, and filtering) are described in the Supporting Information.

Population Genetics Analyses and Scans for Selection

Allele frequency, diversity, and differentiation estimation

To avoid biases in allele frequency and population genetics estimates introduced by NGS (Harismendy et al. 2009; Obbard et al. 2009; Gompert et al. 2010) and a pooling strategy, we implemented maximum-likelihood (ML) estimates of allele frequency (see Supporting Information for methodological details) from read counts, which are shown in Table S4B.

From these imputed allele frequencies, we computed estimates of gene diversity and differentiation over all populations, and also separately for each pair of populations (*Lotus-Medicago* [LM]; *Lotus-Trifolium* [LT]; *Medicago-Trifolium* [MT]). Genetic differentiation among aphid host races was examined using F_{ST} estimates (Wright 1951) calculated according to Weir and Cockerham (1984). We chose the overall heterozygosity across samples as the expected heterozygosity H_e in multiple population estimations. Difference in locus-specific F_{ST} among gene classes (*Or, Gr, Obp, Csp*, nonchemosensory genes) was tested using the nonparametric Kruskal–Wallis rank sum test (Sokal and Rohlf 1981).

F_{ST}-based scans for selection

We initially explored the data for potential outlier SNPs by identifying loci above the 95th and 99th percentiles of the empirical F_{ST} distribution. F_{ST} values were plotted against SNP genomic scaffold positions using release 2.0 of the pea aphid genome (http://www.ncbi.nlm.nih.gov/nuccore/298479576). Conformity to expectation under neutral evolution was statistically tested for all SNPs through a model-based test directed toward the detection of outlier loci. To identify SNP loci showing unusually high F_{ST} values, we approximated the expected distribution of F_{ST} conditional upon heterozygosity for all SNP loci using a modified version of the software package Dfdist (http://sapc34.rdg.ac.uk/~mab/software.html) (Beaumont and Nichols 1996). Since the software package Dfdist was specifically designed for the analysis of biallelic dominant markers (e.g.,

Bonin et al. 2006), we modified it to simulate codominant, biallelic data (code available upon request) (Segurel et al. 2010; Ayala et al. 2011).

We performed 500,000 coalescent simulations of biallelic markers in three samples of eight diploid individuals to characterize the joint distribution of F_{ST} and heterozygosity, using a 10-demes finite-island model and $\theta = 2nN\mu = 0.1$ (where $n = 10$ is the number of demes of size N, and μ is the mutation rate). This particular value of the parameter θ was chosen to match the observed overall gene diversity of the pooled sample, using $\theta = H_e / (1 - H_e)$. We fixed a maximum frequency of the most common allele at 0.999. We used the overall heterozygosity of the pooled sample, which makes the conditional density behave better, particularly for biallelic loci. Outliers were determined by comparing observed distributions with the neutral expectations at the 99% confidence level. Differences in outlier distribution among gene types were tested using log-likelihood ratio tests.

Genes were ranked according to the Poisson probability of the observed or a greater number of SNP outliers, given the number of SNPs in the gene and the overall proportion of outliers (for both the global and the pairwise comparisons), and we then applied a cutoff of $P < 0.05$ and at least three outlier SNPs to categorize genes as outliers.

To test for a potential correlation of F_{ST} level among linked loci, we correlated F_{ST} between pairs of SNPs in distance categories up to 600 kb (500 bp intervals to 10 kb, 1 kb intervals 10–50 kb, 10 kb intervals thereafter, to even out the numbers of SNP pairs per interval) and tested how quickly the correlation falls off with genomic distance in base pairs, assuming an exponential decline and weighting points by the square root of the number of SNP pairs contributing to each correlation (FITCURVE procedure in GenStat).

Synonymous and nonsynonymous diversity

We used the ratio of nonsynonymous to synonymous diversities (π_a/π_s) within and between samples as a proxy for adaptive divergence. The π_a/π_s ratio depends on the constraint on the protein, that is, what proportion of amino acid substitutions is deleterious. Within populations, a locus under less than average constraint would have higher π_a/π_s ratio, because the less intense purifying selection results in greater polymorphism for mildly deleterious nonsynonymous substitutions. We expect the within-race ratio to be similar to the between-race ratio under neutrality, or to be greater in the presence of mildly deleterious alleles because they are unlikely to persist long enough to be shared between races. Positive selection causing divergence between races will elevate the between-race ratio relative to the within-race ratio. π_a/π_s ratios within and between host races were calculated as follows: heterozygosities within (π_{within}) and between ($\pi_{between}$) host races were obtained from allele frequency ML estimates at each nonsynonymous and synonymous SNP; gene wise $\pi_{a(within)}$, $\pi_{a(between)}$, $\pi_{s(within)}$, and $\pi_{s(between)}$ were calculated by averaging over multiple SNPs per gene and neutrality index (*NI*) was defined as the ratio of $\pi_a/\pi_{s\ (between)}$ over $\pi_a/\pi_{s\ (within)}$. We used permutation methods to test whether *NI* was significantly greater or less than 1, $NI > 1$ being suggestive of positive selection on nonsynonymous substitutions (1000 permutations of SNPs within each locus between the synonymous and nonsynonymous classes, recording the number of times that permuted *NI* was less than or greater than the observed *NI*, excluding loci with < 4 synonymous or < 2 nonsynonymous SNPs). We then tested the effect of gene type on *NI*. We also compared *NI* between outlier and nonoutlier genes to test whether outlier genes have different *NI* values and a different *NI* distribution. As the distribution of *NI* departed from normality, we used nonparametric tests (see section Results).

All of the statistical analyses were conducted using the R software environment for statistical computing version 2.13.0 (R Foundation for Statistical Computing, 2011) or GenStat Release 14.1 (VSN International Ltd., 2011).

Results

SEQUENCING AND CAPTURE METRICS

As expected with the capture technology used here (NimbleGen capture arrays), around 54% of all captured and sequenced DNA fragments uniquely mapped to targets (capture specificity, Fig. S1 and Table S3), with a mean read coverage per sample of 58× (Table S3). Importantly, capture specificity was similar across all samples (Fig. S1). Moreover, around 98% of all target loci and exons could be captured (capture uniformity, Fig. S2): among the 203 targeted genes (Table S1), 198 were successfully captured and resequenced, among which 197 were polymorphic (Table S2). Three *Gr* genes could not be detected in the captured pool of sequences (*Gr70, Gr72, Gr74*) and no variation (SNP called from raw alignments) was detected in *Gr64*.

Pairwise population comparisons (*Lotus-Medicago*: LM; *Lotus-Trifolium*: LT; *Medicago-Trifolium*: MT) showed that gene content varies among samples: in particular, we found an absence of some *Or* genes or parts of *Or* genes in the *Lotus* sample (*Or12,* parts of *Or57, Or61, Or74, Or75*), which may indicate either a capture deficiency due to high divergence between sample and reference sequences (*Medicago* origin) or copy-number variation (CNV) among populations in this *Or* multigene family.

After SNP filtering (see Supporting Information), mean depth of coverage per sample was 120-fold, and there remained a total of 9889 biallelic SNPs identified across all three populations (2950 synonymous and 6939 nonsynonymous SNPs), representing 172 target genes and including all gene types (Tables S2 and S4). Diversity and differentiation analyses were performed on this good-quality SNP dataset.

DIVERSITY AND DIFFERENTIATION

The average expected heterozygosity (H_e) at ~10,000 SNP loci across the three host races was 0.126, with some differences among host races (*Lotus*: $H_e = 0.108 \pm 0.001$; *Medicago*: $H_e = 0.120 \pm 0.002$; *Trifolium*: $H_e = 0.118 \pm 0.002$; Kruskal–Wallis chi-squared = 98.28, df = 2, $P < 0.0001$). Gene classes vary little in H_e or in overall proportion of nonsynonymous SNPs (Table S4A).

There was considerable heterogeneity in the single-locus-specific F_{ST} values, ranging from -0.066 to 1 (Table S4B). The multilocus F_{ST} estimate ($F_{ST} = 0.062$) indicates moderate levels of overall genetic differentiation. Pairwise population analysis showed lower differentiation between the *Medicago* and *Trifolium* populations compared with the other population pairs (multilocus F_{ST} [LM] = 0.084; F_{ST} [LT] = 0.083; F_{ST} [MT] = 0.019), as expected from previous work (Ferrari et al. 2008; Peccoud et al. 2009; Ferrari et al. 2011).

F_{ST} distributions differed among gene types (*Or, Gr, Obp, Csp*, nonchemosensory genes) (Fig. S3) and the mean locus-specific F_{ST} estimate significantly differed among gene types (F_{ST} [nonchemosensory genes] = -0.018 ± 0.013; F_{ST} [*Or*] = 0.002 ± 0.013; F_{ST} [*Gr*] = 0.011 ± 0.013; F_{ST} [*Obp*] = -0.012 ± 0.017; F_{ST} [*Csp*] = 0.017 ± 0.013; Kruskal–Wallis chi-squared = 142.713, df = 4, $P < 0.0001$).

DETECTION OF LOCI UNDER DIVERGENT SELECTION

A first indication of SNP loci exhibiting the strongest differentiation is given by the overall F_{ST} distribution, which highlights SNPs above the 95th and 99th percentiles (Fig. S4). Due to the preliminary stage of the pea aphid genome assembly, most target genes occur on different scaffolds, but some genes can be positioned on the same genomic scaffold (Fig. S4). In particular, we observed several clusters of physically linked genes showing high levels of differentiation (Fig. S4; e.g., *Or73-Or62, Gr8-Gr9-Gr10, Gr22-Gr31-Gr20, Or51-Or7-Or15*).

We used the multilocus F_{ST} estimate obtained from all target SNP loci ($F_{ST} = 0.062$, see above) to perform coalescent simulations. When plotted against their respective H_e value (Fig. 1), F_{ST} estimates obtained from most SNPs mapped within the 99% confidence envelope of F_{ST} estimates expected under neutrality. However, some SNPs revealed highly significant departure from neutral expectations, with F_{ST} estimates lying outside the 99% confidence envelope (Fig. 1 and Table S5A). Under global outlier tests using all three populations, signatures of divergent selection were detected at 128 SNP loci (1.3%) while 49 (0.5%), 71 (0.7%), and 35 (0.35%) outlier SNPs were identified for the LM, LT, and MT pairwise population comparisons, respectively (Fig. 1).

The 128 outlier SNPs detected across all three populations were classified according to the type of gene they belong to: 83 outlier SNPs were in *Or* genes (65% of all outlier SNPs), 38 in *Gr* genes (30%), seven in nonchemosensory genes (5%). No outlier loci could be identified in *Obp* and *Csp* genes (Table S5A). There was a significant difference in the incidence of outlier SNPs among genes classified according to their function: the proportion of outlier SNPs in nonchemosensory genes (0.45%) was significantly lower than the proportion of outlier SNPs in *Or* (1.59%) and *Gr* (1.29%) genes (*G*-test: log-likelihood ratio $G = 14.61$, df = 2, $P = 0.0007$), and this pattern was also present in the pairwise comparisons (LM: $G = 6.371$, df = 2, $P = 0.0414$ [proportion significantly higher in *Or* as compared to *Gr* and nonchemosensory genes]; LT: $G = 30.476$, df = 2, $P < 0.0001$ [proportion significantly lower in nonchemosensory genes as compared to *Gr* and *Or*]; MT: $G = 6.852$, df = 2, $P = 0.0325$ [proportion significantly lower in nonchemosensory genes as compared to *Gr* and *Or*]).

Each population pair showed signatures of divergent selection at several SNP loci (Tables S5B–D). However, the *Medicago–Trifolium* pair showed the lowest number of outliers (Table S5D). This is the least-differentiated population pair overall, which could indicate that the separation between these two host plant races was the most recent event of ecological adaptation among those three host races or that the hosts present the most similar environments and so the weakest divergent selection. The low number of outliers is more consistent with the latter interpretation. Results from 16 sampled haplotypes per population suggest that most divergent SNPs segregate within each population, but a few SNPs seem to be alternatively fixed in diverging populations (SNPs in *Gr33, Gr59, Or21, Or6* genes) (Table S5).

How many of these outlier SNPs impact on the protein sequence of the corresponding genes can be determined from synonymous and nonsynonymous SNP assignment. Results indicate that 74 outlier SNPs in the global analysis are nonsynonymous substitutions (Table S5) while 89.8 ([6939 × 128] / 9889) would be expected on the basis of the proportions among all SNPs. These nonsynonymous outlier substitutions might be key changes that may affect the function and/or structure of the corresponding chemoreceptors.

Although an individual nucleotide change can potentially affect host acceptance phenotype, genes showing signs of differentiation at multiple positions may reflect the strongest signature of divergent selection, and thus be considered as the most promising candidate genes for host plant adaptation and speciation in aphids. Figures 2 and 3 highlight those genes that have been categorized as outlier genes on the basis of the low probability of their showing the observed high proportion of outlier SNPs, given their total number of SNPs, and of the presence of three or more outlier SNPs in their sequence (see section Material and methods and Table S6). These results point again at the class of *Or* genes as the most prone to diverge among host races, although a subset

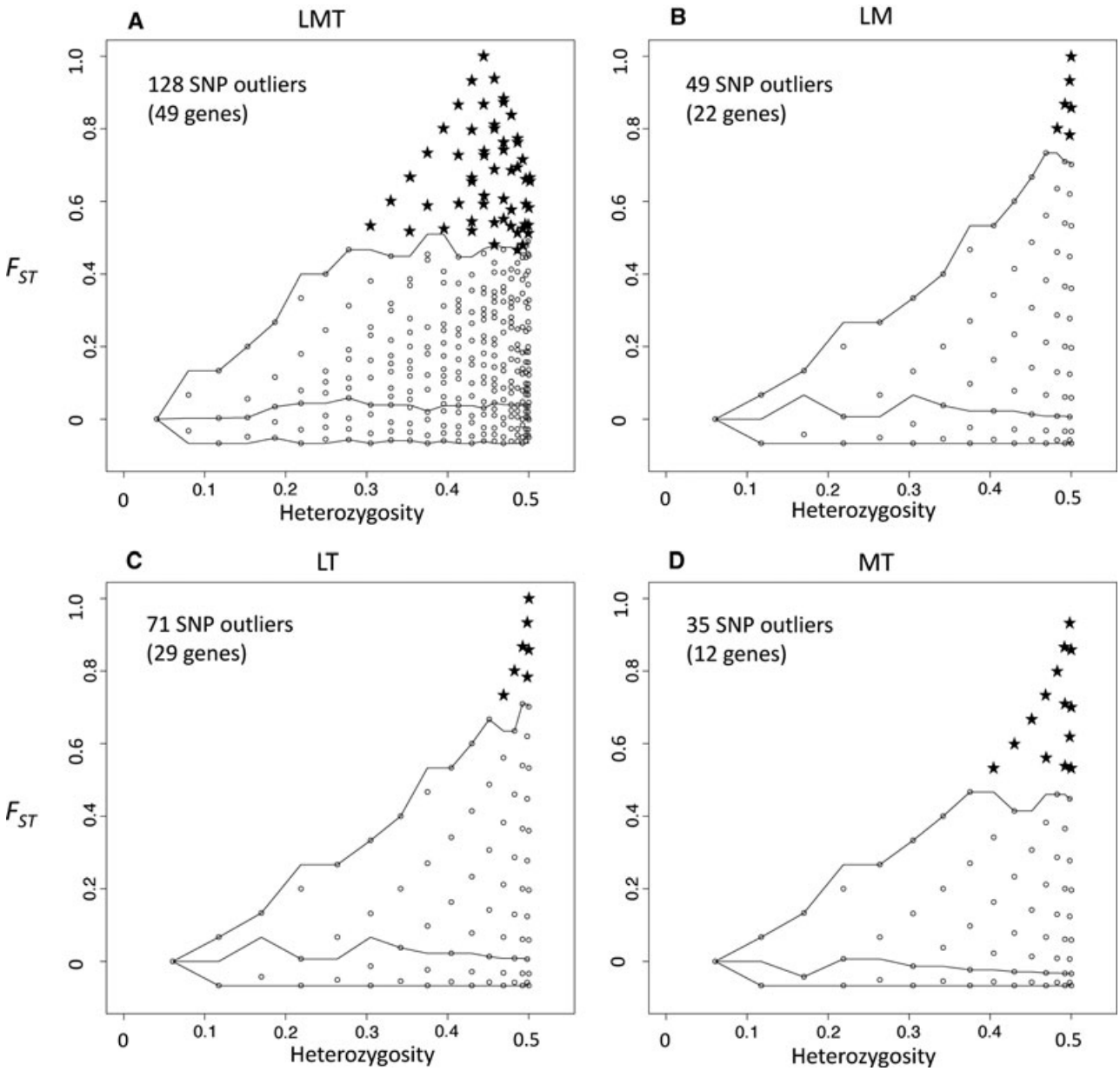

Figure 1. Results of F_{ST}-based scans for selection. Results of Dfdist analyses—each plot illustrates the joint distribution of F_{ST} versus H_e estimates, based on maximum-likelihood imputation of allele counts at 9889 SNP loci (black dots) for (A) the global comparison among *Lotus-Medicago-Trifolium* host races (LMT), (B) the *Lotus-Medicago* host race comparison (LM), (C) the *Lotus-Trifolium* host race comparison (LT), (D) the *Medicago-Trifolium* host race comparison (MT). The zone between the upper and lower black lines represents the neutral expectation at the 99% confidence level. The middle line represents the median of this distribution. The set of outliers revealed by our analyses (see the main text) is indicated in black stars.

of *Gr* genes also seems to contribute to host race divergence. One nonchemosensory gene, *Rad51C*, appeared as an outlier gene, with a very low overall level of polymorphism (Fig. 2). Although most outlier genes were detected in the global analysis, the pairwise population comparisons pointed at outlier genes specific to some pairs of host races and to some additional loci not detected in the global analysis (*Or13, Gr59, Or56*) (Fig. 3). Some of the most divergent genes were among those clustered on the same genome scaffolds (*Or15-Or51; Or62-Or73; Gr8-Gr9-Gr10; Gr20-Gr22-Gr31*, Fig. 2).

The correlation between F_{ST} values of pairs of SNPs in different distance classes declined with distance between the SNPs within scaffolds. An exponential decline explained 25.1% of the variance among correlation coefficients (Fig. S5). The correlation was close to zero for separations greater than 50 kb. Hitchhiking effects around selected loci predict that the correlation would

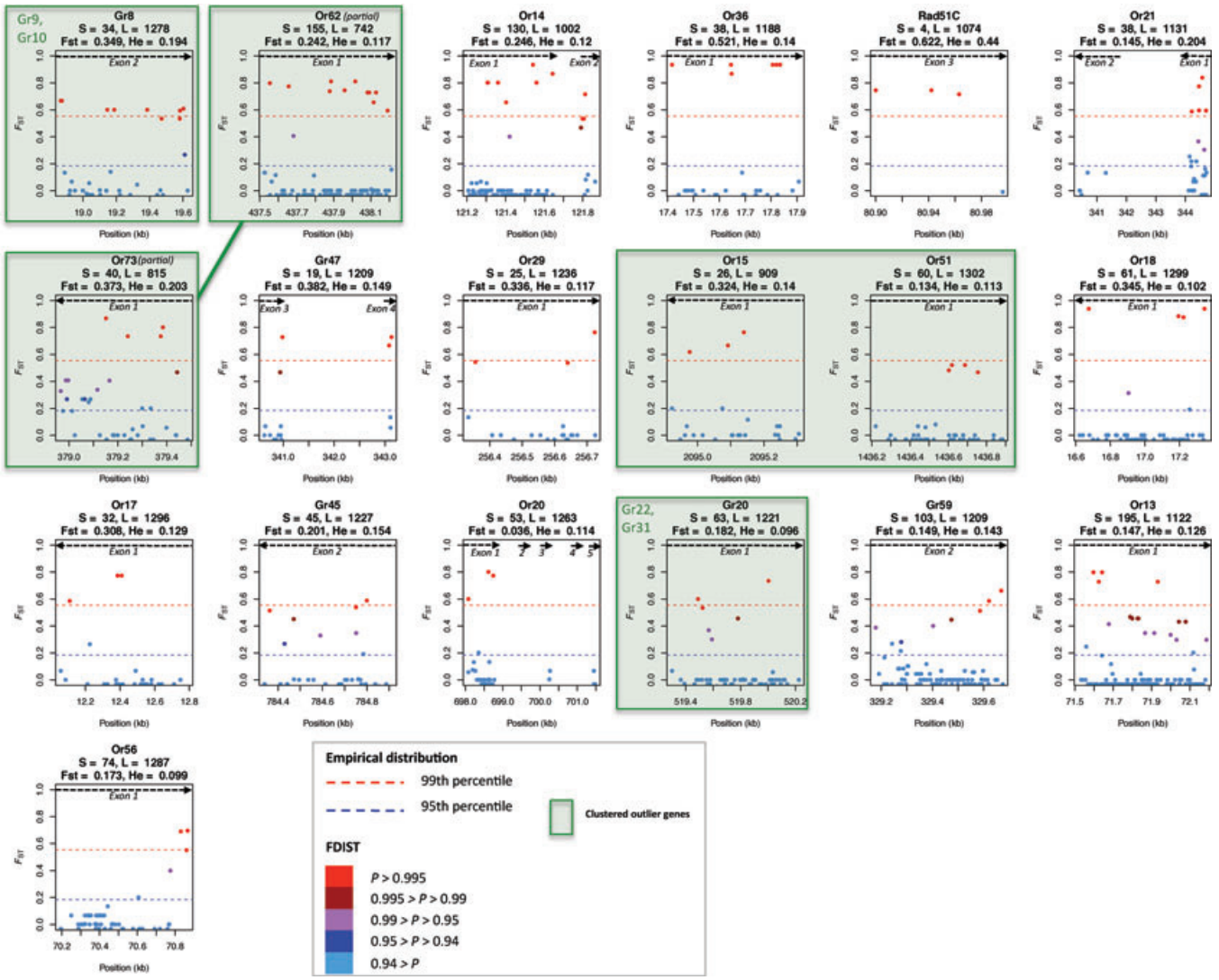

Figure 2. Distribution of F_{ST} along each outlier gene. Each plot represents SNP-based F_{ST} estimates along the sequence (position on scaffolds) of each outlier gene (global analysis). The dot color code indicates the outlier status of each SNP after Dfdist analysis (see legend). Only genes with a significantly greater observed number of SNP outliers than expected given their total number of SNPs and having at least three Dfdist outlier SNPs in their sequence are plotted. Horizontal lines indicate the 99th percentile (red dashed line) and 95th percentile (blue dashed line) of the overall empirical F_{ST} distribution. Above each gene-specific plot are specified the number of segregating sites (*S*), the length of the gene (*L*), the multilocus F_{ST}, and the mean expected heterozygosity (H_e). Green rectangles enclose outlier genes that are clustered on the same genomic scaffold.

decline more slowly on scaffolds containing outlier loci than on other scaffolds. We observed a tendency in this direction but the rates of decline did not differ significantly (fitting separate exponential declines: $F_{2,112} = 1.04$). This suggests that, if hitchhiking does occur around selected genes, its effect is small and it only extends over short distances.

SYNONYMOUS AND NONSYNONYMOUS DIVERSITY

Figure S6 presents the ratios of nonsynonymous to synonymous diversities within ($\pi_a/\pi_{s\ within}$) and between ($\pi_a/\pi_{s\ between}$) host races, for each target gene. The neutrality index *NI* (defined as $[\pi_a/\pi_{s\ between}] / [\pi_a/\pi_{s\ within}]$) was on average significantly lower than 1 (one-sample Wilcoxon rank sum test: mean $NI = 0.984 \pm 0.013$; $V = 2979$; $P < 0.0001$; Fig. S7), suggesting overall evidence for purifying selection. Although some genes showed relatively high *NI* values (Fig. S7), we did not reveal any genes with *NI* significantly greater than 1 (tested by permutation). Gene type did not have any significant effect on median *NI* (Kruskal–Wallis test with *Csp* and *Obp* classes excluded because there were insufficient loci with *NI* estimates, $H = 2.489$; df $= 2$; $P = 0.29$; Fig. S7). Moreover, F_{ST} outlier genes (see section above) did not have significantly different *NI* on average than nonoutlier genes (two-sample Wilcoxon rank sum test: $n_{outlier} = 18$; $NI_{outlier} = 0.989 \pm 0.059$; $n_{nonoutlier} = 133$; $NI_{nonoutlier} = 0.983 \pm 0.013$;

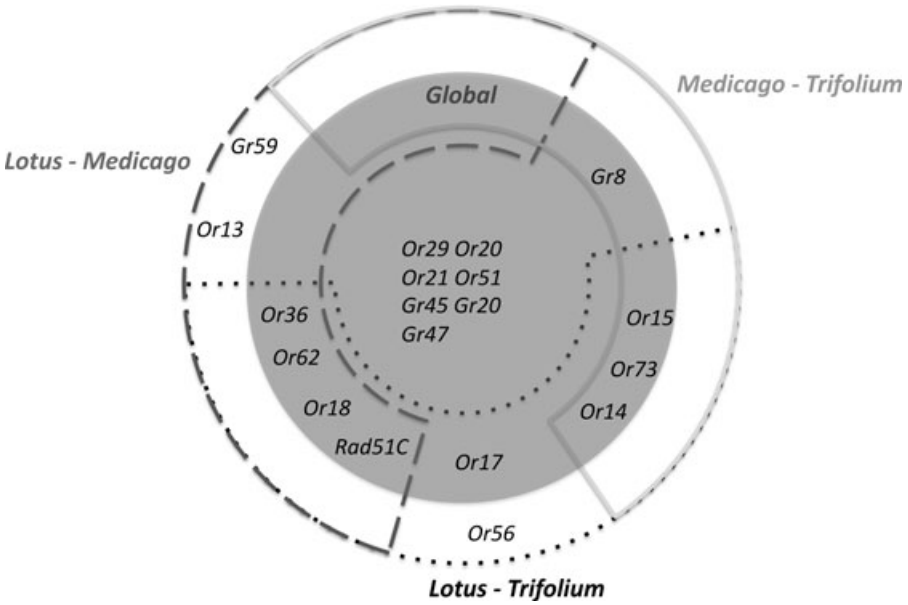

Figure 3. Diagram summarizing common and unique sets of outlier loci among the different pairwise population comparisons and the global comparison. The three pairwise population comparisons (*Lotus-Medicago, Lotus-Trifolium, Medicago-Trifolium*) are represented by the curved areas (defined by the dotted black line; dashed dark gray line; solid light gray line). The gray circle represents the global analysis (*Lotus-Medicago-Trifolium*). The figure highlights the sets of outlier loci identified in each analysis, their specificities, and their overlaps. Genes outside the gray circle are outliers only detected by one or two pairwise comparisons but not in the global comparison. Genes in the center of the gray circle are outliers only detected in the global comparison and not in any pairwise analysis. Finally, genes detected in both pairwise and global analyses are represented in the overlapping zones between the gray circle and the curved areas.

$W = 1236.5$; $P = 0.82$; Fig. S7). The same result was found when comparing genes that contain outlier SNPs and genes that do not contain outlier SNPs (global F_{ST}-based scan for selection) (two-sample Wilcoxon rank sum test: $n_{outlier} = 47$; $NI_{outlier} = 1.004 \pm 0.041$; $n_{nonoutlier} = 104$; $NI_{nonoutlier} = 0.975 \pm 0.005$; $W = 2642$; $P = 0.4274$; Fig. S7). This suggests no strong evidence for positive selection on nonsynonymous substitutions in outlier genes. However, Figure S7 also suggested an unexpected trend: a significant difference in the distribution of *NI* between outlier and nonoutlier genes (Kolmogorov–Smirnov test, $D = 0.247$, $P = 0.027$), with outlier genes showing extreme high and low *NI* values. This indicates that outlier genes tended to have either synonymous or nonsynonymous outlier SNPs, rather than a mixture of the two.

Discussion

Scans for selection are being intensively used to identify the genomic targets of natural selection and loci underlying adaptive traits (Nielsen 2005; Storz 2005; Stinchcombe and Hoekstra 2008). Typically, studies have conducted genome scans for selection using a relatively small number of markers, randomly chosen in the genome, which cannot guarantee that loci involved in adaptation are included (Excoffier et al. 2009). At present, whole-genome or whole-transcriptome scans for selection are being developed to tackle the genetics of adaptation at a larger scale (domestication in plants or animals: Rubin et al. 2010; e.g., adaptation to soil type in *Arabidopsis*: Turner et al. 2010; Lam et al. 2011) and they are becoming increasingly popular in research on the genetics of speciation (Anderson et al. 2010; Elmer et al. 2010; Galindo et al. 2010; Lawniczak et al. 2010; Renaut et al. 2010). Here, we developed an alternative approach, screening a large number of candidate genes whose functions suggest their relevance for ecological specialization and speciation. This candidate gene approach makes it more straightforward to link genotype and phenotype and so interpret the selective pressures acting on these genomic regions (Shikano et al. 2010; Shimada et al. 2011). For the first time we combined the candidate gene approach with sequence capture technology (Burbano et al. 2010; Shen et al. 2011) to address the genetic basis of adaptation and reproductive isolation allows scaling up the classical candidate gene approach to numerous functionally relevant genes and in multigene families.

The main results of our study can be summarized as follows:

(1) We identified a handful of loci of known functional class showing strong signatures of divergence under selection among three aphid host races (Fig. 2). The moderate heterozygosity in all three populations does not suggest any effect of severe bottlenecks on this pattern of differentiation. Similarly, variation in mutation rates is unlikely to explain variation among loci in levels of genetic differentiation because it is not predicted to strongly affect

genetic differentiation for any class of loci when gene flow between populations occurs (Beaumont and Nichols 1996; Hedrick 2005), which is the case among pea aphid host races (Peccoud et al. 2009). Thus, the significantly higher differentiation of multiple SNPs for these loci strongly suggests that gene flow is much more limited at these loci than at other loci in the genome and that they are exposed to strong divergent selection. Previous studies have pointed out that using overall divergence (global F_{ST}) as an expectation may not lead to detection of all highly differentiated loci (Vitalis et al. 2001), but the additional use in our study of pairwise population comparisons should ensure the detection of candidates that are under selection in only a portion of populations (Fig. 3) and thus a good sensitivity of our analysis (Nosil et al. 2008). Moreover, the detection of outliers in this study is quite conservative as far as false positives are concerned: we used the Dfdist program which has a low false positive rate (Beaumont and Nichols 1996) and we only used the 99% envelope to further reduce type I error; in addition, we only retained genes as good candidates if they had a significantly greater number of SNP outliers than expected, given their total number of SNPs and at least three Dfdist outlier SNPs in their sequence. Together with the fact that we evidenced a strong nonrandom distribution of outlier SNPs across gene families (see below), this makes the presence of false positives unlikely for our final set of outlier genes. Even though only a small proportion of SNPs were identified as outliers (1.3% outlier SNPs), the pattern of multiple outlier SNPs in a handful of genes represents a strong signature of selection influencing those genes.

(2) OR and GR are indicated as the key gene families for host plant specialization in the pea aphid complex, suggesting for the first time a role of some chemoreceptor genes in local adaptation and ecologically based speciation. Other chemosensory genes such as binding proteins (OBP, CSP) were not significantly differentiated among aphid host races, but we may have lacked some power to detect outlier SNPs in these very short genes (Table S4). Nonchemosensory genes were, as expected, less affected by divergent selection than chemosensory genes: *Rad51C*, the only nonchemosensory outlier, encodes a strand transfer protein involved in recombinational DNA repair and meiotic recombination. No straightforward functional link with host plant specialization is apparent and so this locus may be influenced by nearby selected loci.

Although functional analysis of outlier genes is still needed and the other potential for other outlier loci needs to be explored, the approach we have adopted here, measuring diversity and differentiation at a large set of loci, targeted by function, has revealed an unprecedentedly detailed picture of the genomic divergence among pea aphid races specialized on different host plants and has identified a group of loci potentially critical to the partial reproductive isolation between these races.

GENETIC ARCHITECTURE OF LOCAL ADAPTATION AND SPECIATION

One major challenge in the field of research on speciation is to gain insights into the genetic architecture of adaptive and reproductive isolating traits and into the dynamics of genomic differentiation during a speciation event. In the pea aphid, our results suggest a polygenic basis for host plant specialization and reproductive isolation, several candidate genes showing some signature of divergent selection. These results echo previous findings obtained from QTL mapping between two North American host races specialized on red clover and alfalfa, which suggested the existence of several complexes of QTLs affecting the key traits for host plant specialization (Hawthorne and Via 2001).

Although at present we have access only to a preliminary assembly of the pea aphid genome, we can show that some genes contain several outlier SNPs (Fig. 2) and that some of the highly differentiated genes identified by our study are physically linked on the same genomic scaffold (Figs. 2 and S4). The influence of selection on these physically linked SNPs and genes might be crucial in the evolution of host plant specialization and reproductive isolation. A current challenge resides in understanding whether they represent independent targets of selection or if these patterns reflect genetic hitchhiking from nearby regions (other genes, regulatory regions) (Maynard Smith and Haigh 1974), and in estimating the size of these highly differentiated genomic regions (Rogers and Bernatchez 2007; Smadja et al. 2008; Via and West 2008; Nosil et al. 2009). A recent theoretical study predicts that adaptation with migration should tend to result in concentrated genetic architectures with divergence of fewer and more tightly linked alleles of larger effect (Yeaman and Whitlock 2011). In aphids, a combined QTL and genome scan approach suggested individual targets of selection could influence neutral markers up to 20 cM away, which would constitute a particularly important effect of genetic hitchhiking around loci involved in local adaptation (Via and West 2008). Here, we show that the level of differentiation is on average correlated among SNPs within 50 kb of one another. Beyond this, the correlation is close to zero both in scaffolds that do not contain any outlier genes and in those that do (Fig. S5). This result suggests the effect of hitchhiking around selected loci is highly localized, on a scale very much smaller than previously suggested for the pea aphid (Via and West 2008) and more in line with that observed recently in *Heliconius* butterflies (Nadeau et al. 2012).

However, some relatively distant pairs of genes (e.g., *Or51-Or15* 600 kb apart) show high levels of differentiation and we cannot rule out the possibility that this pattern is due to more extensive hitchhiking, or divergence hitchhiking (Via and West 2008), of neutral markers in linkage disequilibrium with loci under particularly strong selection. A new linkage and QTL map of key adaptive traits for host plant specialization in the pea aphid

is about to be released (Via, pers. comm.), and together with improved versions of the genome assembly, this will allow for a more complete analysis of genetic linkage among chemosensory genes and of patterns of differentiation along aphid chromosomes.

WHAT TYPES OF GENETIC CHANGES ARE RESPONSIBLE FOR ADAPTATION AND REPRODUCTIVE ISOLATION?

The molecular origin of adaptive changes is a long-term debate: do adaptation and reproductive isolation originate from regulatory or protein evolution (Hoekstra and Coyne 2007; Pavey et al. 2010)? Insights on this question can come from patterns of differentiation at nonsynonymous and synonymous sites along the sequence of outlier genes. Interestingly, we showed that F_{ST} outlier genes tend to have either synonymous or nonsynonymous outlier SNPs, rather than a mixture of the two. Outlier genes having highly differentiated nonsynonymous substitutions might contain functionally significant amino acid changes and thus reveal an impact of selection on the protein function among diverging populations. Little is currently known of the protein features of chemoreceptors in aphids and thus interesting perspectives lie in future proteomics and functional analysis of those candidate molecules. In contrast, some outlier genes showed an inflated differentiation restricted to or largely affecting multiple synonymous SNPs. This observation suggests that the divergent selection acting on these outlier loci does not act at the structural level but rather points to expression differences as the targets of selection. In turn, this would imply that selection acts on the control regions of those genes, differentiation in the coding region resulting from selective sweeps in the neighboring sequences (Barton 2000). If this sweep signature is due to a recent adaptation, it is expected that a *cis*-regulatory change outside of the coding region would have caused it (Cai et al. 2009). This suggests the possibility that expression divergence, in addition to protein change, can drive adaptive changes at chemosensory genes in the pea aphid, which may or may not be linked to CNV in these multigene families (e.g., Itsara et al. 2009; Bigham et al. 2010; Scavetta and Tautz 2010). Our capture experiment suggests some variation in gene content between *Lotus* and the two other host races (*Or12,* parts of *Or57, Or61, Or74, Or75* missing in the *Lotus* population), which might indicate CNV in this gene family. However, specific analysis of CNV at chemoreceptor loci will be needed to explore this hypothesis.

OR AND GR GENES AS KEY LOCI FOR HOST PLANT ADAPTATION AND SPECIATION IN APHIDS

Among the four chemosensory gene families targeted in this study as potential candidates for host plant adaptation and speciation in the pea aphid, two gene families were detected to have high level of differentiation among host races: the OR and the GR gene families. Our previous work on the molecular evolution of those chemoreceptor genes in the pea aphid genome had revealed patterns of evolution under positive selection in some of the most recently duplicated genes in these families (some *Gr* and *Or* clades, see Smadja et al. 2009), a result consistent with the potentially strong divergent evolutionary pressures experienced by aphids when entering new niches during host shifts or host specialization events. Interestingly, F_{ST} outlier genes among the three host races revealed in the present study tend to fall among the most recently duplicated members of these gene families (Smadja et al. 2009). Molecular and electrophysiology studies had already suggested a role for peripheral chemoreception in host plant use in other phytophagous insects (*Rhagoletis pomonella*: Olsson et al. 2006b; McBride 2007; *Drosophila sechellia*: McBride et al. 2007), but this is the first time, to our knowledge, that population genomics has been applied to the whole gene superfamily (*Gr* + *Or*) to tackle divergence among host races and test for the role of chemoreceptor genes in host plant specialization and speciation.

Aphids do not identify their host plant from a distance. Indeed, winged forms of the pea aphid reject alternate hosts shortly after a brief first penetration of the plant tissues (Caillaud and Via 2000), following the typical behavior of aphids on nonhost plants (Powell et al. 2006). In contrast, host acceptance takes longer, the key step being the probing of epidermal and subepidermal plant tissues with their antennae and stylet (Caillaud and Via 2000; Margaritopoulos et al. 2005). Aphids seem to distinguish their host through cues located on the plant surface or in subcutaneous tissues perceived prior to the initiation of feeding, but key host cues are surprisingly not known and we do not know if these cues act as attractants or repellents as in other phytophagous insects (e.g., habitat avoidance in *R. pomonella*: Forbes et al. 2005; plant volatile attraction in *Argyresthia conjugella* Bengtsson et al. 2006; Feder and Forbes 2007). Nutrition on alternate hosts can be artificially triggered by epidermal extracts from the home plant (Del Campo et al. 2003), suggesting that olfactory cues at the surface of the plant can be sufficient. The F_{ST} outlier OR genes identified in our study may be involved in this early stage of host acceptance. In addition, superficial penetration of plant tissues seems to follow exploration of the plant surface and thus intracellular metabolites detected via gustation of ingested epidermal or subepidermal cell contents may inhibit the take-off reflex, which otherwise follows probing (Caillaud and Via 2000; Powell et al. 2006). Outlier *Gr* genes are likely to be involved in this step of host acceptance, which may be even more crucial in the early stage of host race formation (see *Medicago* results). No sugar receptor (putative sugar receptors based on orthology with other insect species: Gr1–6, Smadja et al. 2009) was identified among the outlier *Gr* genes detected in our study, and this is consistent with behavioral results, which indicate that aphids do not need to reach phloem sap to accept a plant as a host (Caillaud and Via 2000).

Response to chemical cues in insects is usually a complex mechanism: it commonly requires the combined activation of several chemoreceptors, chemical signals varying in the number of receptors they activate (Hallem et al. 2004; Dahanukar et al. 2005; Rutzler and Zwiebel 2005; Hallem and Carlson 2006). In our system, we do not know if the different OR and GR under divergent selection act separately or in combination in response to one major cue or a complex blend of plant cues. However, protein or expression divergence at those chemoreceptor loci can lead to changes in receptor specificity and/or receptor neuron sensitivity (Olsson et al. 2006a,c, 2009), which could underlie preference for different host plants among the *Lotus, Medicago*, and *Trifolium* host races. This does not rule out possible additional differences in the central integration of those chemical signals.

PLEIOTROPIC EFFECTS OF CHEMORECEPTORS IN SPECIATION

The identification of some *Or* and *Gr* genes as highly divergent loci among the three studied host races has some interesting consequences for possible scenarios of speciation in the pea aphid. The most direct way chemoreceptors can promote speciation is by their impact on assortative mating, as a by-product of chemical recognition of plant cues, host plant acceptance (or avoidance) behavior, and reproduction on hosts (Via 1999). Interestingly, chemoreceptor genes can have multiple other effects on reproductive isolation. Results from a QTL study on North American populations suggested close physical linkage or pleiotropy among genes underlying host performance and host acceptance (Hawthorne and Via 2001). Such pleiotropy would suggest that host acceptance and feeding behaviors of winged pea aphids depend, at least partly, on the same loci. This hypothesis is consistent with our results identifying *Or* and *Gr* loci as strongly differentiated genes among aphid host races as these genes could be involved not only in the chemical recognition of plant cues but also in the recognition of food stimulants, which are known to influence nutrition and parturition (Powell et al. 2006), and thus fitness on specific hosts in aphids. It is thus plausible that the same set of GRs or closely linked ones is involved in the recognition of both host acceptance cues and food stimulants. This would be a very favorable scenario for speciation, as pleiotropy would automatically generate a correlation between host acceptance and host performance traits, which will not be impeded by gene flow (Smadja and Butlin 2011). Another possible level of pleiotropy concerns the link between ecological isolation (driven by host acceptance and host adaptation) and sexual isolation. Recent studies on other insect species suggest a role of chemoreceptors, and in particular ORs, in the reception of sex pheromones and thus mate choice (*Ostrinia nubilalis*: Lassance et al. 2011; *Heliothis*: Vasquez et al. 2011). Sexual isolation has not been assessed among pea aphid host races yet, but it is possible that it contributes to reproductive isolation. If so, the same loci, some chemoreceptor genes, could underlie several major components of reproductive isolation involved in local adaptation and sexual isolation and evolve under both natural and sexual selection, representing key drivers of speciation in this system and a very favorable scenario for the evolution of barriers to gene flow.

CONCLUSION AND PERSPECTIVES

This study illustrates how the large-scale candidate gene approach we developed here is an efficient way to narrow down loci potentially involved in adaptation and speciation. It should complement whole-genome scan and QTL mapping studies, and one interesting perspective in the aphid system would be to test for the colocation of outlier chemoreceptor loci with QTL for host acceptance and performance that will soon become available. We only start to envisage how capture technologies might contribute to scaling up population genetics studies but this experiment clearly highlights the efficacy of capture technologies to scale up candidate gene analysis while generating sequence information. This route becomes essential when dealing with multigene families. Our approach, by revealing a detailed picture of the genomic divergence among pea aphid host races that are partially reproductively isolated, paves the way for a promising alternative route to the genetics of adaptation and speciation.

In the near future, it will be important to build on these results by analyzing differentiation in larger samples and by exploring divergence in more dimensions. We may want to extend our study of key genetic changes to expression and CNV, to characterize all of the possible dimensions of divergence at chemosensory loci. Moreover, the pea aphid system offers an exceptional opportunity to further address differentiation at chemosensory candidate loci along the continuum of differentiation among different aphid biotypes and species (Peccoud et al. 2009), which should help us to gain insights into the dynamics of differentiation during a speciation process. In parallel, functional analysis of key loci will provide the means to link molecular and phenotype variation.

ACKNOWLEDGMENTS

We thank C. Peichel, associate editor, and two anonymous referees for their useful comments on this manuscript. We thank NBAF-Liverpool, and in particular Dr. L. Olohan for her advice for DNA sample preparation and for running the capture experiment. We thank Dr. O. Kaiser, Senior Manager Systems Support at Roche Diagnostics Limited for his support in the array design. We would also like to thank D. Obbard, J. Welch, R. Kofler, R. Nichols, and N. Bierne for fruitful discussions on data analysis. CS acknowledges the European Commission for funding (Grant "Aphid Genomics" RU112509, Marie Curie Fellowship, CORDIS Actions, Sixth Framework Programme FP6) and the Centre National pour la Recherche Scientifique (CNRS). RV acknowledges partial funding by the Agence Nationale de la Recherche programme BLANC "EMILE" 09-BLAN-0145-01. JJZ acknowledges the grant-aided support from the Biotechnology and Biological Sciences Research Council of the United

Kingdom. RKB acknowledges funding from NERC. We thank the European Commission network Molecular Adaptation in Ecologically Relevant Organisms for funding. This is publication ISEM no. 2012-014.

LITERATURE CITED

Anderson, J. B., J. Funt, D. A. Thompson, S. Prabhu, A. Socha, C. Sirjusingh, J. R. Dettman, L. Parreiras, D. S. Guttman, A. Regev, et al. 2010. Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. Curr. Biol. 20:1383–1388.

Ayala, D., M. C. Fontaine, A. Cohuet, D. Fontenille, R. Vitalis, and F. Simard. 2011. Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. Mol. Biol. Evol. 28:745–758.

Barton, N. H. 2000. Genetic hitchhiking. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 355:1553–1562.

Baxter, S. W., S. E. Johnston, and C. D. Jiggins. 2009. Butterfly speciation and the distribution of gene effect sizes fixed during adaptation. Heredity 102:57–65.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. Ser. B Biol. Sci. 263:1619–1626.

Bengtsson, M., G. Jaastad, G. Knudsen, S. Kobro, A. C. Backman, E. Pettersson, and P. Witzgall. 2006. Plant volatiles mediate attraction to host and non-host plant in apple fruit moth, *Argyresthia conjugella*. Entomologia Experimentalis Et Applicata 118:77–85.

Bigham, A., M. Bauchet, D. Pinto, X. Y. Mao, J. M. Akey, R. Mei, S. W. Scherer, C. G. Julian, M. J. Wilson, D. L. Herraez, et al. 2010. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet. 6:14.

Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.

Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Mol. Biol. Evol. 23: 773–783.

Burbano, H. A., E. Hodges, R. E. Green, A. W. Briggs, J. Krause, M. Meyer, J. M. Good, T. Maricic, P. L. F. Johnson, Z. Y. Xuan, et al. 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. Science 328:723–725.

Butlin, R. K., J. Galindo, and J. W. Grahame. 2008. Review. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:2997–3007.

Butlin, R. K., J. R. Bridle, and D. Schluter. 2009. Speciation and patterns of diversity. Cambridge Univ. Press, Cambridge, U.K.

Cai, J. J., J. M. Macpherson, G. Sella, and D. A. Petrov. 2009. Pervasive hitchhiking at coding and regulatory sites in humans. PLoS Genet. 5:13.

Caillaud, M. C., and S. Via. 2000. Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. Am. Nat. 156:606–621.

Consortium, T. I. A. G. 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8:e1000313.

Croset, V., R. Rytz, S. F. Cummins, A. Budd, D. Brawand, H. Kaessmann, T. J. Gibson, and R. Benton. 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. 6:e1001064.

Dahanukar, A., E. A. Hallem, and J. R. Carlson. 2005. Insect chemoreception. Curr. Opin. Neurobiol. 15:423–430.

Del Campo, M. L., S. Via, and M. C. Caillaud. 2003. Recognition of host-specific chemical stimulants in two sympatric host races of the pea aphid *Acyrthosiphon pisum*. Ecol. Entomol. 28:405–412.

Dulac, C., and A. T. Torello. 2003. Molecular detection of pheromone signals in mammals: from genes to behaviour. Nat. Rev. Neurosci. 4:551–562.

Dworkin, I., and C. D. Jones. 2009. Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. Genetics 181:721–736.

Elias, M., Z. Gompert, C. Jiggins, and K. Willmott. 2008. Mutualistic interactions drive ecological niche convergence in a diverse butterfly community. PLoS Biol. 6:2642–2649.

Elmer, K. R., S. Fan, H. M. Gunter, J. C. Jones, S. Boekhoff, S. Kuraku, and A. Meyer. 2010. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Mol. Ecol. 19: 197–211.

Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103:285–298.

Feder, J. L., and A. A. Forbes. 2007. Habitat avoidance and speciation for phytophagous insect specialists. Funct. Ecol. 21:585–597.

Felsenstein, J. 1981. Skepticism towards Santa Rosali, or why are there so few kinds of animals? Evolution 35:124–138.

Ferrari, J., H. C. J. Godfray, A. S. Faulconbridge, K. Prior, and S. Via. 2006. Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. Evolution 60:1574–1584.

Ferrari, J., S. Via, and H. C. J. Godfray. 2008. Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. Evolution 62:2508–2524.

Ferrari, J., J. A. West, S. Via, and H. C. J. Godfray. 2011. Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. Evolution: Early View 66:375–390.

Forbes, A. A., J. Fisher, and J. L. Feder. 2005. Habitat avoidance: overlooking an important aspect of host-specific mating and sympatric speciation? Evolution 59:1552–1559.

Frantz, A., M. Plantegenest, L. Mieuzet, and J. C. Simon. 2006. Ecological specialization correlates with genotypic differentiation in sympatric host-populations of the pea aphid. J. Evol. Biol. 19:392–401.

Galindo, J., J. W. Grahame, and R. K. Butlin. 2010. An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. J. Evol. Biol. 23:2004–2016.

Gardiner, A., D. Barker, R. K. Butlin, W. C. Jordan, and M. G. Ritchie. 2008. *Drosophila* chemoreceptor gene evolution: selection, specialisation and genome size. Mol. Ecol. 17:1648–1657.

Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton, NJ.

Gompert, Z., M. L. Forister, J. A. Fordyce, C. C. Nice, R. J. Williamson, and C. A. Buerkle. 2010. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of Lycaeides butterflies. Mol. Ecol. 19:2455–2473.

Haas, F., M. A. Pointer, N. Saino, A. Brodin, N. I. Mundy, and B. Hansson. 2009. An analysis of population genetic differentiation and genotype-phenotype association across the hybrid zone of carrion and hooded crows using microsatellites and MC1R. Mol. Ecol. 18:294–305.

Hallem, E. A., and J. R. Carlson. 2006. Coding of odors by a receptor repertoire. Cell 125:143–160.

Hallem, E. A., M. G. Ho, and J. R. Carlson. 2004. The molecular basis of odor coding in the *Drosophila* antenna. Cell 117:965–979.

Hallem, E. A., A. Dahanukar, and J. R. Carlson. 2006. Insect odor and taste receptors. Annu. Rev. Entomol. 51:113–135.

Harismendy, O., P. Ng, R. Strausberg, X. Wang, T. Stockwell, K. Beeson, N. Schork, S. Murray, E. Topol, S. Levy, et al. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol. 10:R32.

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

Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:1633–1638.

Hodges, E., Z. Xuan, V. Balija, M. Kramer, M. N. Molla, S. W. Smith, C. M. Middle, M. J. Rodesch, T. J. Albert, G. J. Hannon, et al. 2007. Genome-wide in situ exon capture for selective resequencing. Nat. Genet. 39:1522–1527.

Hoekstra, H. E., and J. A. Coyne. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution 61:995–1016.

Itsara, A., G. M. Cooper, C. Baker, S. Girirajan, J. Li, D. Absher, R. M. Krauss, R. M. Myers, P. M. Ridker, D. I. Chasman, et al. 2009. Population analysis of large copy number variants and hotspots of human genetic disease. Am. J. Hum. Genet. 84:148–161.

Kent, L. B., and H. M. Robertson. 2009. Molecular evolution of the sugar receptors in insects. BMC Evol. Biol. 9:41.

Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection: models and experiments. Am. Nat. 159:21–35.

Kronforst, M. R., L. G. Young, D. D. Kapan, C. McNeely, R. J. O'Neill, and L. E. Gilbert. 2006. Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. Proc. Natl. Acad. Sci.USA 103:6575–6580.

Lam, H. M., X. Xu, X. Liu, W. B. Chen, G. H. Yang, F. L. Wong, M. W. Li, W. M. He, N. Qin, B. Wang, et al. 2011. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 43:387–387.

Lassance, J. M., S. M. Bogdanowicz, K. W. Wanner, C. Lofstedt, and R. G. Harrison. 2011. Gene genealogies reveal differentiation at sex pheromone olfactory receptor loci in pheromone strains of the European corn borer, *Ostrinia nubilalis*. Evolution 65:1583–1593.

Lawniczak, M. K. N., S. J. Emrich, A. K. Holloway, A. P. Regier, M. Olson, B. White, S. Redmond, L. Fulton, E. Appelbaum, J. Godfrey, et al. 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. Science 330:512–514.

Lexer, C., Z. Lai, and L. H. Rieseberg. 2004. Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. New Phytol. 161:225–233.

Lowry, D. B., and J. H. Willis. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. 8:e1000500.

Margaritopoulos, J. T., C. Tsourapas, M. Tzortzi, O. M. Kanavaki, and J. A. Tsitsipis. 2005. Host selection by winged colonisers within the *Myzus persicae* group: a contribution towards understanding host specialisation. Ecol. Entomol. 30:406–418.

Matsuo, T. 2008. Genes for host-plant selection in *Drosophila*. J. Neurogenet. 22:195–210.

Matsuo, T., S. Sugaya, J. Yasukawa, T. Aigaki, and Y. Fuyama. 2007. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. PLoS Biol. 5:985–996.

Maynard Smith, J., and J. Haigh. 1974. The hitchhiking effect of a favourable gene. Genet. Res. 23:23–35.

McBride, C. S. 2007. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proc. Natl. Acad. Sci. USA 104:4996–5001.

McBride, C. S., J. R. Arguello, and B. C. O'Meara. 2007. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics 177:1395–1416.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, R. H. ffrench-Constant, M. L. Blaxter, et al. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367:343–353.

Nei, M., Y. Niimura, and M. Nozawa. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat. Rev. Genet. 9:951–963.

Nielsen, R. 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39:197–218.

Niimura, Y. 2009. On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. Genome Biol. Evol. 1:34–44.

Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26:160–167.

Nosil, P., S. R. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316–336.

Nosil, P., D. J. Funk, and D. Ortiz-barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Obbard, D. J., J. J. Welch, K.-W. Kim, and F. M. Jiggins. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. PLoS Genet. 5:e1000698.

Olsson, S., C. E. Linn, and W. L. Roelofs. 2006a. The chemosensory basis for behavioral divergence involved in sympatric host shifts. I. Characterizing olfactory receptor neuron classes responding to key host volatiles. J.Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 192:279–288.

Olsson, S. B., C. E. Linn, A. Michel, H. R. Dambroski, S. H. Berlocher, J. L. Feder, and W. L. Roelofs. 2006b. Receptor expression and sympatric speciation: unique olfactory receptor neuron responses in F-1 hybrid *Rhagoletis* populations. J. Exp. Biol. 209:3729–3741.

Olsson, S. B., C. E. Linn, and W. L. Roelofs. 2006c. The chemosensory basis for behavioral divergence involved in sympatric host shifts II: olfactory receptor neuron sensitivity and temporal firing pattern to individual key host volatiles. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 192:289–300.

Olsson, S. B., C. E. Linn, J. L. Feder, A. Michel, H. R. Dambroski, S. H. Berlocher, and W. L. Roelofs. 2009. Comparing peripheral olfactory coding with host preference in the *Rhagoletis* species complex. Chem. Senses 34:37–48.

Pavey, S. A., H. Collin, P. Nosil, and S. M. Rogers. 2010. The role of gene expression in ecological speciation. Pp. 110–129. Year in evolutionary biology. Blackwell Publishing, Oxford, U.K.

Peccoud, J., A. Ollivier, M. Plantegenest, and J. C. Simon. 2009. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. Proc. Natl. Acad. Sci. USA 106:7495–7500.

Peccoud, J., and J.-C. Simon. 2010. The pea aphid complex as a model of ecological speciation. Ecol. Entomol. 35:119–130.

Powell, G., C. R. Tosh, and J. Hardie. 2006. Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. Annu. Rev. Entomol. 51:309–330.

Renaut, S., A. W. Nolte, and L. Bernatchez. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Mol. Ecol. 19:115–131.

Robertson, H. M., J. Gadau, and K. W. Wanner. 2010. The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. Insect Mol. Biol. 19:121–136.

Rogers, S. M., and L. Bernatchez. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonas* sp Salmonidae) species pairs. Mol. Biol. Evol. 24:1423–1438.

Rubin, C. J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, et al. 2010.

Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587-U145.

Rutzler, M., and L. J. Zwiebel. 2005. Molecular biology of insect olfaction: recent progress and conceptual models. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 191:777–790.

Sanchez-Gracia, A., F. G. Vieira, and J. Rozas. 2009. Molecular evolution of the major chemosensory gene families in insects. Heredity 103:208–216.

Scavetta, R. J., and D. Tautz. 2010. Copy number changes of CNV regions in intersubspecific crosses of the house mouse. Mol. Biol. Evol. 27:1845–1856.

Schymura, D., M. Forstner, A. Schultze, T. Krober, L. Swevers, K. Iatrou, and J. Krieger. 2010. Antennal expression pattern of two olfactory receptors and an odorant binding protein implicated in host odor detection by the malaria vector *Anopheles gambiae*. Int. J. Biol. Sci. 6:614–626.

Segurel, L., S. Lafosse, E. Heyer, and R. Vitalis. 2010. Frequency of the AGT Pro11Leu polymorphism in humans: does diet matter? Annals Hum. Genet. 74:57–64.

Servedio, M. R. 2009. The role of linkage disequilibrium in the evolution of premating isolation. Heredity 102:51–56.

Servedio, M. R., and M. A. F. Noor. 2003. The role of reinforcement in speciation: theory and data. Annu. Rev. Ecol. Syst. 34:339–364.

Shaw, K. L., and S. C. Lesnick. 2009. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. Proc. Natl. Acad. Sci. USA 106:9737–9742.

Shen, P. D., W. Y. Wang, S. Krishnakumar, C. Palm, A. K. Chi, G. M. Enns, R. W. Davis, T. P. Speed, M. N. Mindrinos, and C. Scharfe. 2011. High-quality DNA sequence capture of 524 disease candidate genes. Proc. Natl. Acad. Sci. USA 108:6549–6554.

Shikano, T., J. Ramadevi, and J. Merila. 2010. Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). Mol. Biol. Evol. 27:2775–2789.

Shimada, Y., T. Shikano, and J. Merila. 2011. A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. Mol. Biol. Evol. 28:181–193.

Smadja, C., and R. K. Butlin. 2009. On the scent of speciation: the chemoreception system and its role in premating isolation. Heredity 102: 77–97.

———. 2011. A framework for comparing processes of speciation in the presence of gene flow. Mol. Ecol. 20:5123–5140.

Smadja, C., J. Galindo, and R. K. Butlin. 2008. Hitching a lift on the road to speciation. Mol. Ecol. 17:4177–4180.

Smadja, C., P. Shi, R. K. Butlin, and H. M. Robertson. 2009. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrthosiphon pisum*. Mol. Biol. Evol. 26: 2073–2086.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. The principles and practices of statistics in biological research. Freeman, W.H. and Company, New York.

Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–170.

Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14:671–688.

Touhara, K., and L. B. Vosshall. 2009. Sensing odorants and pheromones with chemosensory receptors. Annu. Rev. Physiol. 71:307–332.

Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat. Genet. 42:260-U242.

Vasquez, G. M., P. Fischer, C. M. Grozinger, and F. Gould. 2011. Differential expression of odorant receptor genes involved in the sexual isolation of two Heliothis moths. Insect Mol. Biol. 20:115–124.

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.

———. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. Evolution 53:1446–1457.

———. 2009. Natural selection in action during speciation. Proc. Natl. Acad. Sci. USA 106:9939–9946.

Via, S., and J. West. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17:4334–4345.

Via, S., A. C. Bouck, and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. Evolution 54:1626–1637.

Vitalis, R., K. Dawson, and P. Boursot. 2001. Interpretation of variation across marker loci as evidence of selection. Genetics 158:1811–1823.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population-structure. Evolution 38:1358–1370.

Wright, S. 1951. The genetical structure of populations. Ann Hum Genet. 15:323–354.

Wu, C.-I. 2001. The genetic view of the process of speciation. J. Evol. Biol. 14:851–865.

Yeaman, S., and M. C. Whitlock. 2011. The genetic architecture of adaptation under migration—selection balance. Evolution 65:1897–1911.

Zhou, J.-J., F. G. Vieira, X.-L. He, C. Smadja, R. Liu, J. Rozas, and L. M. Field. 2010. Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrthosiphon pisum*. Insect Mol. Biol. 19:113–122.

Zhu, M. J., and S. H. Zhao. 2007. Candidate gene identification approach: progress and challenges. Int. J. Biol. Sci. 3:420–427.

Associate Editor: C. Peichel

Supporting information

The following supporting information is available for this article:

Figure S1. Capture specificity.
Figure S2. Capture uniformity.
Figure S3. F_{ST} distributions.
Figure S4. Distribution of F_{ST} by position on scaffolds.
Figure S5. Correlation between F_{ST} values at different SNPs against the distance between SNPs.
Figure S6. Comparison of the ratio π_a/π_s between host races and the ratio π_a/π_s within host races.
Figure S7. Neutrality index (*NI*) for outlier and nonoutlier genes.
Table S1. Target genes.
Table S2. Number and type of genes under study: from initial to analyzed targets.
Table S3. Read numbers and mapping.
Table S4. All SNPs.
Table S5A. Outlier SNPs—global analysis.
Table S6. Outlier genes probability = Poisson probability of the observed or a greater number of SNP outliers given the number of SNPs in the gene and the overall proportion of outliers in red: $P < 0.05$ in gray shade: genes with three or more outlier SNPs in their sequence.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.