Genetica (2011) 139:565–573
DOI 10.1007/s10709-010-9514-4

SI - GOS

# Genetic hitchhiking associated with life history divergence and colonization of North America in the European corn borer moth

**Erik B. Dopman**

Received: 29 April 2010 / Accepted: 24 October 2010 / Published online: 21 November 2010


**Abstract** A primary goal for evolutionary biology is to reveal the genetic basis for adaptive evolution and reproductive isolation. Using Z and E pheromone strains the European corn borer (ECB) moth, I address this problem through multilocus analyses of DNA polymorphism. I find that the locus *Triose phosphate isomerase* (*Tpi*) is a statistically significant outlier in coalescent simulations of demographic histories of population divergence, including strict allopatric isolation, restricted migration, secondary contact, and population growth or decline. This result corroborates a previous QTL study that identified the *Tpi* chromosomal region as a repository for gene(s) contributing to divergence in life history. Patterns of nucleotide polymorphism at *Tpi* suggest a recent selective sweep and genetic hitchhiking associated with colonization of North America from Europe ~200 generations ago. These results indicate that gene genealogies initially diverge during speciation because of selective sweeps, but differential introgression may play a role in the maintenance of differentiation for sympatric populations.

**Keywords** Divergence · Speciation · Selection · Gene flow · Reproductive isolation · Polymorphism · *Ostrinia nubilalis* · European corn borer

**Electronic supplementary material** The online version of this article (doi:10.1007/s10709-010-9514-4) contains supplementary material, which is available to authorized users.

E. B. Dopman (✉)
Department of Biology, Tufts University, 165 Packard Avenue, Medford, MA 02155, USA
e-mail: erik.dopman@tufts.edu

## Introduction

Quantitative-trait locus (QTL) mapping and multilocus analyses of DNA polymorphism across populations are two complementary approaches for locating genes or chromosomal regions contributing to reproductive isolation or adaptive evolution. Perhaps the most straightforward of the two approaches is QTL mapping, which has enabled the characterization of dozens of unexplored genomes, as well as the genetic architectures of many compelling examples of phenotypic evolution (reviewed in Erickson et al. 2004). In contrast to QTL mapping, multilocus analyses do not require prior knowledge of character differences. In this framework, loci that are outliers with respect to within-population polymorphism or between-population divergence represent candidate genome regions that have been targets of recent adaptive evolution (e.g., Harr et al. 2002; reviewed in Storz 2005). Because many genes contributing to reproductive barriers will have been shaped by selection (reviewed by Coyne and Orr 2004), at least in principle, scans for outlier loci between diverged populations are detecting both adaptive evolution and reproductive isolation.

In recent years, genetic mapping and multilocus polymorphism methods have been integrated to find loci for ecologically-relevant characters (reviewed in Stinchcombe and Hoekstra 2008). Both were recently used for partially isolated Z and E pheromone strains of the European corn borer moth (ECB) (Dopman et al. 2004, 2005). In these studies, genealogical exclusivity between pheromone strains was strongest at *Triose phosphate isomerase* (*Tpi*), a gene that was tightly linked to a QTL for life history differences and temporal isolation. This result suggests that QTL and "outlier mapping" provide complementary estimates for the locations of genes contributing to

diversification. However, there is an interpretational challenge for multilocus analyses of DNA polymorphism because neutral demographic changes associated with population divergence or speciation (e.g., substructure or bottlenecks) can produce locus-specific outlier patterns, such as genealogical exclusivity, by random chance (Przeworski 2002; Thornton and Andolfatto 2006). In these situations, the potential for false positives can be high.

By taking advantage of the availability of both QTL and multilocus data for the ECB, I revisit the data of Dopman et al. (2004, 2005) and take an explicit population genetics perspective to assess whether statistically supported outlier loci exist among gene regions. I then incorporate new data from "ancestral" European populations to obtain an historical context for patterns of descent at *Tpi*.

Given population-genetic support for the hypothesis that selection has shaped the *Tpi* genome region, a final issue that is confronted concerns the type of selection that has operated. Dopman et al. (2005) proposed two hypotheses to account for locus-specific exclusivity in the ECB. One scenario involves a selective sweep of a beneficial allele(s) contributing to adaptive trait divergence between Z and E populations, followed by genetic hitchhiking in the *Tpi* gene region (Maynard Smith and Haigh 1974). A second possibility is differential introgression, which describes recurrent introgression of, and negative selection against, maladaptive alleles (Barton and Hewitt 1981; Charlesworth et al. 1997).

The distinction between selective sweeps and differential introgression as explanations for genealogical divergence is important for at least two reasons. First, it may help account for variation in the size of differentiated regions between diverged populations or species (e.g., Harr 2006; Emelianov et al. 2004; Via and West 2008). As noted by Via and West (2008), theoretical simulations suggest that large genome intervals of differentiation can result from differential introgression (>1 cM; Charlesworth et al. 1997), whereas genetic hitchhiking will typically yield more localized signatures (5–10 kb) (Charlesworth et al. 1997; Kim and Stephan 2002). Second, distinguishing between these models can help predict whether genome intervals containing phenotypic targets of selection have been historically important as reproductive barriers (Schluter 2000). For selective sweeps, reproductive isolation may have only recently evolved as a by-products of adaptation, depending on whether trait differences cause pre or postzygotic isolation. In contrast, during differential introgression, selection has directly restricted gene flow between sympatric populations at those chromosome regions containing QTL for trait differences.

## The European corn borer

The ECB is native to Europe, North Africa, and Western Asia, but its current range includes all corn-producing areas (*Zea mays*) of the United States and Canada east of the Rocky Mountains (Mutuura and Monroe 1970). ECB colonized North America at multiple sites by infesting imported broomcorn (*Sorghum vulgare*) from Italy and Hungary early in the 20th century (Smith 1920; Caffrey and Worthley 1927). ECB larvae feed on both wild and agricultural plants, but are commonly found on maize (Caffrey and Worthley 1927). In North America and across most of its range, the ECB consists of a series of sympatric, partially interfertile populations (Klun and Cooperators 1975; Anglade et al. 1984). Incomplete prezygotic barriers prevent gene exchange among ECB populations (Dopman et al. 2010), but beyond these trait differences, moths are difficult to distinguish (Liebherr 1974). Indeed, besides *Tpi*, pheromone strains share extensive genetic polymorphism across their genome (see Dopman et al. 2005 and Malausa et al. 2007).

The dominant form of isolation in the ECB stems from the use of different sexual communication systems (Dopman et al. 2010). In the Z strain, females produce and males preferentially respond to a 3:97 mixture of (*E*) and (*Z*)-11-tetradecenyl acetate, whereas in the E strain females produce and males preferentially respond to a 99:1 (*E*)/(*Z*) blend (Roelofs et al. 1987). Behavioral isolation arises from a single major autosomal gene contributing to pheromone blend (*Pher*) differences (Lassance et al. 2010), and a second major sex-linked factor contributing to differences in male response (*Resp*) (Roelofs et al. 1987; Dopman et al. 2004). Population surveys reveal that at many sites in Europe and North America only Z strain males and females occur, that Z and E strain moths occur sympatrically at a number of sites, and that the E strain rarely occurs alone (Klun and Cooperators 1975; Kochansky et al. 1975; Cardé et al. 1978; Anglade et al. 1984). Both pheromone strains occur in New York, and sympatric populations feeding primarily on maize have been documented at a number of localities (Roelofs et al. 1985; Glover et al. 1991). The situation in Europe is more complicated, because there appear to be two distinct "host races". The race feeding on *Artemisia* (mugwort) or *Humulus* (hop) uses the E pheromone, whereas the one feeding primarily on maize uses the Z pheromone (Bourguet et al. 2000; Thomas et al. 2003).

ECB populations are also characterized by variation in number of generations per year (voltinism). In Europe and in the United States, moths are univoltine in the north and bivoltine in the south, but multivoltine populations occur at some southern US sites (Showers et al. 1975). Differences in voltinism patterns were apparent soon after the original

introductions of ECB to the northeastern United States (Caffrey and Worthley 1927). Although life cycle differences can be a consequence of environmental variation, early studies suggested a genetic component and sex-linked inheritance (McLeod 1978; Reed et al. 1981). In New York, ECB larvae diapause over the winter, then pupate and emerge as adults in the spring and summer. Differences in post-diapause development time, the time to pupation under temperature and photoperiod conditions conducive to breaking diapause, in large part determine number of generations per year. Postdiapause development (PDD) time is affected primarily by a major gene (*Pdd*) or genes on the sex chromosome (Glover et al. 1992; Dopman et al. 2005). Because the periods of flight of univoltine and bivoltine moths are displaced in time (Roelofs et al. 1985), genetic differences affecting voltinism can have important consequences for the extent of gene flow in natural populations (Dopman et al. 2010).

## Materials and methods

### Data collection

Nucleotide sequences were collected for five loci, which are located in the mitochondrial genome, on the Z sex chromosome, and on an autosome. All five loci are protein-coding regions, some of which consist of introns and exons. The loci include the Z-linked genes *Tpi*, *Kettin* (*Ket*), and *Lactate dehydrogenase* (*Ldh*), the autosomal locus *Pheromone binding protein* (*Pbp*), and the mitochondrial gene *cytochrome oxidase I* (*COI*). Due to alignment difficulties with the noncoding region in *Ldh*, only coding region was used in analysis. Molecular methods used to collect the sequence data can be found elsewhere (Dopman et al. 2005).

The sequences of the five loci from a maize-feeding North American sample ($n = 17$ bivoltine E-strain moths, $n = 20$ bivoltine and univoltine Z-strain moths) and from an outgroup species ($n = 1$) (Asian corn borer, *Ostrinia furnacalis*) have been previously published (Dopman et al. 2005), as have the sequences for *Tpi* from E-strain mugwort-hop and Z-strain maize host races from France ($n = 28$ distinct haplotypes of bivoltine moths; Malausa et al. 2007). Samples from both geographic regions represent multiple, wide-spread populations ($\geq 5$ populations per region). New *Tpi* sequences from a second European sample of maize-feeding moths from Italy and Hungary are reported here ($n = 9$ bivoltine E-strain, $n = 10$ bivoltine Z-strain), which have been deposited in GenBank under accession numbers XXX-XXX. Italian moths are from a population in Piacenza, and Hungarian moths are from a population in Kety.

### Outlier mapping

I assessed the concordance of outlier and QTL mapping for North American pheromone strains by conducting tests of neutral-coalescent models. Models were based on known recent history, current population biology, and plausible hypothetical scenarios that can generate locus-specific genetic exclusivity. Table 1 describes the classes of models that were used. My goal was not to identify the most probable demographic model, but rather to assess how realistic historical scenarios and their parameter values impact inferences of selection (e.g., Haddrill et al. 2005; Stajich and Hahn 2005). I acknowledge that there may be a model, perhaps more complex than envisioned here, that is more appropriate.

To summarize genealogical patterns of exclusivity, I used $F_{ST}$ from Hudson et al. (1992),

$$F_{ST} = \pi_D / (\pi_D + \pi_S),$$

where $\pi_S$ = mean within-population diversity, and $\pi_D$ = absolute between-population diversity (Charlesworth 1998). As $F_{ST}$ summarizes only one component of DNA diversity, Tajima's $D$ (Tajima 1989) and Fay and Wu's $H$ (Fay and Wu 2000) were used to make a more thorough assessment of model fit, and both statistics were calculated separately for Z and E strains in DnaSP (v. 4.0) (Rozas

**Table 1** General description of neutral models

| Model name | Abbreviation | Description |
|---|---|---|
| Standard neutral model | SNM | A single, randomly mating population of constant size |
| Isolation | I | A single population splits into two daughter populations (strict isolation) |
| Migration | M | Constant migration between two populations |
| Secondary contact | SC | Isolation followed by constant migration between two populations |
| Bottleneck | B | Isolation or migration model with an instantaneous population bottleneck |
| Growth | G | Migration model with instantaneous population growth |
| Isolation with migration[a] | IM | Isolation with migration and population size change |

[a] Isolation with migration model of Hey (2005)

et al. 2003). Multilocus comparisons of polymorphism and divergence between species compared the ECB to the ACB (Hudson et al. 1987), using the program HKA (distributed by Jody Hey).

I follow the approach used in other multilocus studies in which the null model is conditioned on the average Watterson's estimate for the population-mutation parameter ($\theta$) and the sample size (deme 1 = 17 E borers and deme 2 = 20 Z borers; e.g., Haddrill et al. 2005; Stajich and Hahn 2005). I initially evaluated the standard neutral model (SNM), which assumes a single panmictic population of constant size. Tests of the SNM were made without recombination to assess whether any loci were outliers under this assumption, and significance was assessed through 10,000 simulations. Other models were evaluated by generating simulated data sets ($n \geq 1{,}000$) in MS (Hudson 2002).

MS simulations were carried out with recombination (average $C$ across genes excluding *COI*), as the lack of recombination can make neutrality tests overly conservative (Przeworski et al. 2001). By convention, population parameters used in coalescent simulations are relative to the effective population size for a diploid gene. Since the effective copy numbers for Z chromosome and mtDNA are, respectively, 3/4 and 1/4 of that for autosomes, for each model, parameters and time were scaled accordingly. Statistics from the simulated data sets were calculated in MSSTATS (Thornton 2003). Specific details of coalescent models can be found in the Supplementary Methods. Outlier loci were conservatively identified if the probability of observing locus summaries fell below the Bonferroni corrected value of 0.01.

### A ghost of divergence past?

The relative support for a recent or an ancestral origin of genetic divergence at *Tpi* was evaluated through comparisons of DNA sequences of European ECB. Divergence and ancestor-descendent evolutionary patterns were measured by $F_{ST}$ and genealogical reconstruction. *Tpi* genealogy reconstructions were conducted in MEGA4 (Tamura et al. 2007), using the neighbor-joining method (Saitou and Nei 1987) with Tajima-Nei distances (Tajima and Nei 1984). Branch support was evaluated using the bootstrap test (10,000 replicates). All positions containing gaps and missing data were eliminated from the dataset.

## Results

### Outlier mapping

For *Tpi*, $F_{ST}$ between strains from North America was high (0.79) and statistically significantly different ($P \leq 0.01$) from that expected under the SNM (Tables 1, 2). In contrast, at *COI*, *Ket*, *Ldh*, and *Pbp*, $F_{ST}$ was low ($F_{ST} \leq 0.06$) and within SNM expectations. For *Pbp*, $F_{ST}$ was negative, indicating a lack of between-strain diversity. In addition to the unusual levels of nucleotide diversity at *Tpi*, the locus shows atypical patterns with respect to the site-frequency spectrum. With the exception of *Tpi*, values for $D$ and $H$ across loci fall within SNM expectations (Table 3). At *Tpi* $D$ and $H$ yield strongly negative values in the E strain ($D = -2.2$, $P \leq 0.01$; $H = -11.9$, $P \leq 0.01$).

Multi-locus HKA tests using noncoding sites revealed significant heterogeneity among loci (E, $P = 0.03$; Z, $P = 0.02$, nonsyn. and syn. sites were n.s.). *Tpi* was the largest contributor to the total $\chi^2$ statistic, accounting for >56% of the total deviation (*Ket* and *Pbp* were the other two loci with noncoding sites).

Most of the demographic models from Table 1 performed no better than the SNM in accounting for patterns of nucleotide variation among loci. For example, in the strict isolation model (Model I in Table 1), values for divergence time ($T$) that could account for high $F_{ST}$ at *Tpi* (e.g., $P > 0.01$ for $T = 0.5$, 0.75, and 1.87) were not able to account for low $F_{ST}$ at *COI*, *Ket*, *Ldh*, and *Pbp* ($P < 0.01$). The migration (M) and secondary contact (SC) models also performed poorly in that the explored parameter space for migration ($M = 4N_em$) and constant effective population size ($N_e$) could not account for the data ($P < 0.01$). In these models variance in $F_{ST}$ was too narrow to simultaneously account for *Tpi* and the other loci. An example can be seen in Fig. 1, in which observed $F_{ST}$ values for *Tpi*, *Ldh*, and *COI* are plotted against their expected distributions under equilibrium migration ($M = 3$).

In contrast to models in which $N_e$ was constant, the two models that incorporated changes in population size, bottleneck (B) and growth (G) (Table 1), succeeded in explaining $F_{ST}$ values. An acceptable bottleneck (B) model

**Table 2** $F_{ST}$ and components of diversity for Z and E pheromone strains of ECB

| Locus | $F_{ST}$ | $\pi_S$ | $\pi_D$ | $\pi_E/\pi_Z$ |
|---|---|---|---|---|
| *COI* | 0.05 | 0.001 | 0.0001 | 1.43 |
| *Ket* | 0.03 | 0.006 | 0.0002 | 1.24 |
| *Ldh* | 0.02 | 0.015 | 0.0003 | 1.33 |
| *Pbp* | −0.03[a] | 0.021 | −0.0006[a] | 1.04 |
| *Tpi* | **0.79** | 0.001 | 0.0050 | 0.50 |
| *Tpi Ancestor*[b] | 0.07 | 0.003 | 0.0002 | 2.75 |

Bold $F_{ST}$ indicates significance ($P \leq 0.01$) under the standard-neutral model

[a] Negative $F_{ST}$ indicates a lack of between-strain diversity

[b] ECB from Italy and Hungary

**Table 3** Summary statistics and results of neutrality tests for E-strain ECB and Z-strain ECB

| Locus | *N* | Length | $S^a$ | $K^b$ | $C^c$ | $\Theta^d$ | $\pi^e$ | *D* | *H* |
|---|---|---|---|---|---|---|---|---|---|
| E-strain ECB | | | | | | | | | |
| *COI* | 17 | 1,195 | 8 | 0.023 | – | 0.002 | 0.001 | −1.295 | −0.125 |
| *Ket* | 17 | 704 | 17 | 0.011 | 0.001 | 0.007 | 0.007 | −0.159 | 3.772*† |
| *Ldh* | 17 | 192 | 10 | 0.014 | 0.552 | 0.015 | 0.017 | 0.380 | 0.904 |
| *Pbp* | 12 | 474 | 28 | 0.030 | 0.023 | 0.020 | 0.021 | 0.235 | 0.242 |
| *Tpi* | 17 | 1,454 | 11 | 0.021 | 0.000 | 0.002 | 0.001 | **−2.235*†** | **−11.912*** |
| Z-strain ECB | | | | | | | | | |
| *COI* | 20 | 1,195 | 7 | 0.023 | – | 0.002 | 0.001 | −1.543 | −2.642† |
| *Ket* | 20 | 703 | 21 | 0.011 | 0.006 | 0.008 | 0.006 | −1.302 | 2.853*† |
| *Ldh* | 20 | 192 | 8 | 0.012 | 0.144 | 0.012 | 0.013 | 0.288 | 0.158 |
| *Pbp* | 20 | 469 | 40 | 0.031 | 0.036 | 0.025 | 0.020 | −0.681 | −1.621 |
| *Tpi* | 20 | 1,444 | 14 | 0.019 | 0.002 | 0.003 | 0.002 | −1.220 | −6.737* |

Bold values indicate significance ($P \leq 0.01$) under the standard-neutral model. For an growth and bottleneck model that could account for multilocus patterns of $F_{ST}$ in North America ($P \geq 0.01$), significance for *D* and *H* was estimated. * indicates significance ($P \leq 0.01$) under the growth model and † indicates significance ($P \leq 0.01$) under the bottleneck model. See "Results" for details

[a] Number of polymorphic sites

[b] Average percent sequence divergence between ECB and ACB

[c] Per-site recombination rate

[d] Heterozygosity, $4N_e u$, using the number of polymorphic sites

[e] Heterozygosity, $4N_e u$, using the average number of nucleotide differences per site

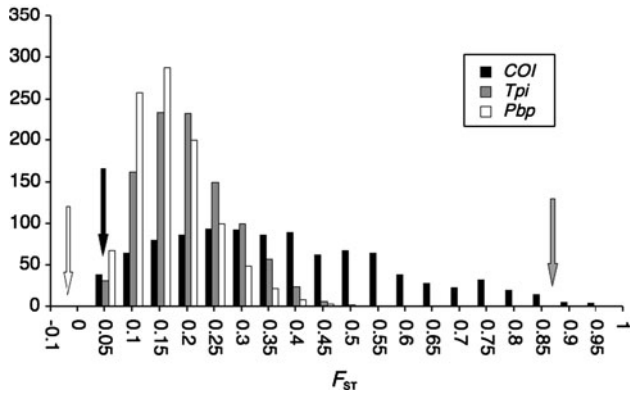


**Fig. 1** Simulated values of $F_{ST}$ (10,000) expected under equilibrium migration ($M = 3$) for *COI* (mtDNA), *Tpi* (Z sex chromosome), and *Pbp* (autosome). Observed $F_{ST}$ values from Table 2 are indicated by *arrows*

was found in which an older, strong bottleneck ($S_b = 2.5$, $t_{b1} = 0.075 \times 4N_e$ generations) occurred in the presence of limited gene flow ($M = 1.0$). Similarly, a recent ($0.00375 \times 4N_e$ generations) 10-fold population growth (G) model with restricted gene exchange ($M = 1.0$) explained $F_{ST}$ for all loci. However, values of *D* and *H* at *Tpi* in the E strain departed significantly from those expected under both null models, as did values at other loci (Table 3). Thus, of the more than 350 strictly neutral demographic histories that were explored, none were consistent with the North American multilocus sample.

## A ghost of divergence past?

Figure 2 depicts the *Tpi* genealogy for the sample of ECB from Europe (France, Italy, and Hungary) and from North America. According to the genealogy, the extreme pattern of genetic differentiation between North American strains did not arise in Europe. A caveat for this and any other analysis that includes presumptive ancestral populations is that ancestral diversity has been sampled. I cannot exclude the possibility that it was not; nevertheless, records indicate that ECB colonized North American from Italy and Hungary (Smith 1920; Caffrey and Worthley 1927). $F_{ST}$ values between Z and E moths across European regions were low (<0.07), a pattern that corroborates a previous analysis of genetic differentiation in France (Malausa et al. 2007). The genealogy suggests that most North American E moths are a derived subset of "ancestral" E-strain diversity at *Tpi*. Unlike the North American E sample, which forms a relatively distinct group, E moths from Europe are widely distributed across the tree. In contrast to the E strain, North American Z moths more clearly retain diversity from Europe. Many of these insects share identical sequences with both Z and E moths from Europe, and as a group they are widely distributed across the genealogy.

The per locus ratio of within-strain diversity for E and Z moths, $\pi_E/\pi_Z$, corrects for differences in mutation rate and recombination rate among loci, and is useful for detecting population and locus-specific changes in

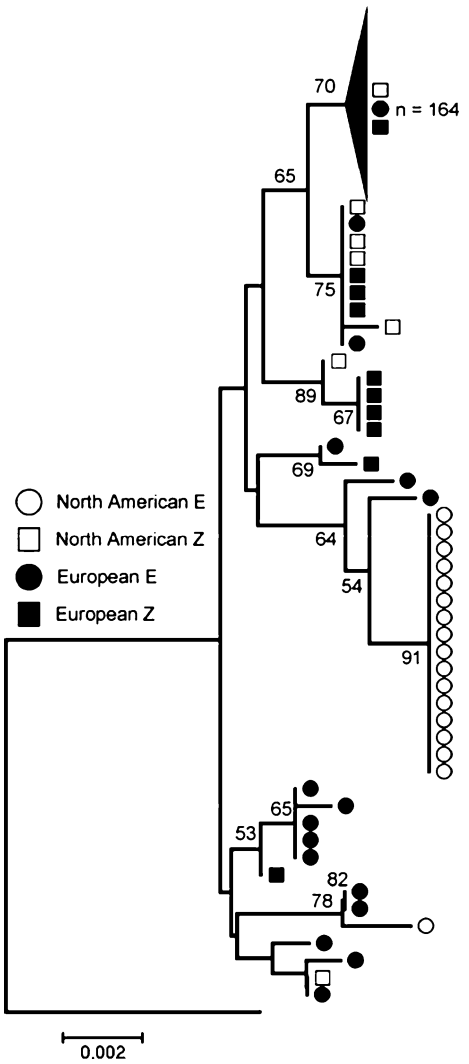


**Fig. 2** The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.05076268 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10,000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 1,132 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4

diversity (e.g., Schlötterer 2002). The $\pi_E/\pi_Z$ ratio is 0.5 at *Tpi* in North America, whereas the mean at other loci is 1.26 (Table 2). These comparisons substantiate the genealogical interpretation that the E strain disproportionately contributes to genetic differentiation at *Tpi* in North America (e.g., via a reduction in $\pi_S$). Indeed, compared with European moths from Italy and Hungary, $\pi_S$ at *Tpi* is reduced by 44% in the North American sample (Table 2). $\pi_D$ is elevated by a factor of 25 in the same comparison (Table 2), suggesting that genetic differentiation at *Tpi* is explained by both a reduction in E strain diversity and an increase in absolute between-strain diversity.

## Discussion

### A test of the outlier method for the functional characterization of genomes

One of most attractive features of outlier methods for identifying genome regions contributing to adaptation or reproductive isolation is that they can be applied to natural populations of just about any species (reviewed in Storz 2005 and Stinchcombe and Hoekstra 2008). Outlier loci are detected by various features of nucleotide polymorphism, such as genetic differentiation between populations, skews in the site-frequency spectrum, or reductions in nucleotide diversity. Random genetic drift produces outliers, but in the case of genetic differentiation, the mitochondrial genome is commonly expected become an outlier first because its reduced effective-copy number relative to nuclear loci (e.g., Hudson and Coyne 2002). In contrast to this outcome, differentiation between North American pheromone strains of ECB is 15 times greater at the Z-linked *Tpi* locus than at the next most differentiated locus, which is, not surprisingly, the mitochondrion (*COI*) (Table 2). In fact, simulation results indicate that locus-specific genetic differentiation at *Tpi* is too great to be explained by many neutral demographic changes associated with population divergence or speciation, including strict isolation, migration, and secondary contact (I, M, and SC in Table 1; see Supplementary Methods). When incorporating change in population size into these models, however, ECB multilocus patterns become more likely.

One consequence of a change in effective population size is an increase in the variance of coalescence time among loci. Such a history presents challenges for outlier mapping because a sample of loci may contain an apparent outlier(s) by chance (e.g., Haddrill et al. 2005; Thornton and Andolfatto 2006), rather than because of natural or sexual selection. Indeed, patterns of DNA polymorphism following recovery from a bottleneck, or after population expansion, often mimic those that are expected after the selective sweep of a beneficial mutation. As North American ECB have a history of colonization, which commonly includes periods of population decline and/or growth, the potential for successfully applying outlier mapping to this species would seem uncertain. Indeed, *Tpi* was not detected as an outlier with respect to genetic differentiation when an old bottleneck or recent expansion was evaluated (model B and G in Table 1). However, these same models were

unable to account for extreme departures in the site-frequency spectrum at *Tpi* (Table 3). Thus, *Tpi* remained an outlier, but only when multiple features of nucleotide polymorphism were considered.

In the large number of realistic demographic scenarios that were investigated (>350), *Tpi* is consistently a statistically significant outlier compared to other loci. Patterns of nucleotide polymorphism from other loci fall within null expectations under the simplest of demographic models, a single panmictic population of constant size (SNM, Table 1), but a significant departure is observed for all statistics at *Tpi* ($F_{ST}$, *D*, and *H*; Tables 2 and 3). Given the modeling results, outlier status at *Tpi* can be attributed to an unspecified selective component. Signatures of selection in the region containing *Tpi* could stem from tight linkage to a gene(s) contributing to a reproductive barrier or a target of adaptive evolution. This conclusion agrees with the independent QTL map for North American ECB strains (Dopman et al. 2004, 2005), in which a major QTL for postdiapause development (*Pdd*) time differences is indistinguishable in map position from *Tpi*. In contrast, the major locus for male response (*Resp*), maps over 20 cM away from *Tpi*/*Pdd*. The concordance of outlier and QTL mapping argues for the reliability of outlier mapping as a stand-alone method for the functional characterization of the ECB genome.

## The nature of selection during incipient speciation

One feature that distinguishes the selective sweep and differential introgression models is the manner in which genetic divergence is achieved. Selective sweeps increase divergence (e.g., $F_{ST}$) at neutral sites between populations by eliminating within-population diversity ($\pi_S$) during the fixation of positively selected alleles (Slatkin and Wiehe 1998). In this process, divergence is expected to be greatest when different rare alleles reach fixation in different populations. However, the maximum *absolute* level of between-population diversity ($\pi_D$) should not exceed the equilibrium level of within-population diversity at neutral loci (Charlesworth et al. 1997; Charlesworth 1998). In contrast, the long-term consequences of the selective removal of maladaptive alleles introduced by introgression can increase $\pi_D$ at linked sites beyond that expected for $\pi_S$ under neutrality (Charlesworth 1998), whereas $\pi_S$ at linked markers is almost completely preserved (Charlesworth et al. 1997).

Limited support for differential introgression is seen when evaluating genetic divergence at *Tpi* in light of the predictions of two models. First, coalescent simulations of various demographic histories (Table 1) indicate that the value of $\pi_D$ at *Tpi* (0.005) is not greater than $\pi_S$ at neutral loci ($P < 0.001$). Second, as stated above, compared to other loci and to *Tpi* in "ancestral" populations, $\pi_S$ at *Tpi* in North America is reduced rather than conserved (Table 2). Finally, results of Charlesworth et al. (1997) suggest that differential introgression will only have a quantifiable effect on diversity over evolutionary time (e.g., $>N_e$ generations), but genetic differentiation followed colonization of North America $\sim$200 generations ago (Fig. 2). If differential introgression were to explain multilocus patterns, signatures of selection would be expected in both descendent and ancestral populations.

Results presented here provide evidence that gene genealogies diverge during speciation because of local adaptation and selective sweeps. However, the maintenance of divergence in sympatry may often require a separate explanation. In the case of the ECB there is a possibility that differential introgression has recently been activated near *Tpi* because of *Pdd*, or another a tightly linked "speciation gene". Hybrid ECB are not uncommon at sympatric sites. If gene flow was on the order of the hybrid proportion (5–15%; Roelofs et al. 1985; Klun and Huettel 1988), then changes in allele frequency would be expected after only 2–3 generations. Yet, a common Z-strain *Tpi* allele has remained absent in NY E-strain populations for over 40 generations (Glover et al. 1991; Dopman et al. 2005). Either selection is actively removing maladapted introgressed alleles in the *Tpi* region through differential introgression, or most hybrids do not backcross and the rate of contemporary gene flow is low at these sites (e.g., $N_e m \approx 1$). Efforts to disentangle these hypotheses will help us to understand the role of selection during the origin, maintenance, and growth of genetic divergence during the speciation process, which ultimately results in daughter populations that form exclusive genetic groups across genome regions.

**Acknowledgments** Rick Harrison provided thoughtful comments on this manuscript, and endless hours of discussion (and argument) on speciation in the many years that preceeded its creation. For Rick's patience, generosity, and wisdom, EBD is immensely grateful. EBD would also like to thank Steve Bogdanowicz, Denis Bourguet, Sergine Ponsard, and Gábor Szócs for providing or helping to procure European ECB. The simulation work was improved by suggestions and assistance from Kevin Thornton. This work was supported in part by a USDA NIFA award (2010-65106-20610).

## References

Anglade P, Stockel J, Cooperators IWGO (1984) Intraspecific sex-pheromone variability in the European corn borer, *Ostrinia nubilalis* Hbn. (Lepidoptera, Pyralidae). Agronomie 4:183–187

Barton NH, Hewitt GM (1981) Hybrid zones and speciation. In: Atchley WR, Woodruff DS (eds) Evolution and speciation. Cambridge University Press, Cambridge, pp 109–145

Bourguet D, Bethenod MT, Trouvé C, Viard F (2000) Host-plant diversity of the European corn borer *Ostrinia nubilalis*: what value for sustainable transgenic insecticidal Bt maize? Proc R Soc B Biol Sci 67:1177–1184

Caffrey DJ, Worthley LH (1927) A progress report on the investigations of the European corn borer. U.S.D.A., Dept. Bull. No. 1476. Government Printing Office, Washington

Cardé RT, Roelofs WL, Harrison RG, Vawter AT, Brussard PF, Mutuura A, Munroe E (1978) European corn borer: pheromone polymorphism or sibling species? Science 199:555–556

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. Mol Biol Evol 15:538–543

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res 70:155–174

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland

Dopman EB, Bogdanowicz SM, Harrison RG (2004) Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). Genetics 167:301–309

Dopman EB, Perez L, Bogdanowicz SM, Harrison RG (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. Proc Natl Acad Sci USA 102:14706–14711

Dopman EB, Robbins PS, Seaman A (2010) Reproductive isolation between Z and E pheromone strains of European corn borer. Evolution 64:881–902

Emelianov I, Marec F, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proc Roy Soc B Biol Sci 271:97–105

Erickson D, Fenster C, Stenoien HK, Price D (2004) Quantitative trait locus analyses and the study of evolutionary process. Mol Ecol 13:2505–2522

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Glover TJ, Knodel JJ, Robbins PS, Eckenrode CJ, Roelofs WL (1991) Gene flow among three races of European corn borers (Lepidoptera:Pyralidae) in New York State. Environ Entomol 20:1356–1362

Glover TJ, Robbins PS, Eckenrode CJ, Roelofs WL (1992) Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. Arch Insect Biochem Physiol 20:107–117

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005) Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Gen Res 15:790–799

Harr B (2006) Genomic islands of differentiation between house mouse subspecies. Gen Res 16:730–737

Harr B, Kauer M, Schlotterer C (2002) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in Drosophila melanogaster. Proc Natl Acad Sci USA 99: 12949–12954

Hey J (2005) On the number of New World Founders: a population genetic portrait of the peopling of the Americas. PLoS Biol 3:e193

Hudson RR (2002) Generating samples under a Wright-fisher neutral model of genetic variation. Bioinformatics 18:337–338

Hudson RR, Coyne J (2002) Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153–159

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589

Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765–777

Klun JA, Cooperators (1975) Insect sex pheromones: intraspecific pheromonal variability of *Ostrinia nubilalis* in North America and Europe. Environ Enomol 4:891–894

Klun JA, Huettel MD (1988) Genetic regulation of sex pheromone production and response: Interaction of sympatric pheromonal types of European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). J Chem Ecol 14:2047–2061

Kochansky J, Cardé RT, Liebherr J, Roelofs WL (1975) Sex pheromones of the European corn borer in New York. J Chem Ecol 1:225–231

Lassance JM, Groot AT, Liénard MA, Antony B, Borgwardt C, Andersson F, Hedenström E, Heckel DG, Löfstedt C (2010) Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. Nature. http://www.dx.doi.org/10.1038/nature09058

Liebherr J (1974) Studies on two strains of European corn borer, Ostrinia nubilalis (Hübner), thesis, Cornell University

Malausa T, Leniaud L, Martin JF et al (2007) Molecular differentiation at nuclear loci in French host races of the European corn borer (*Ostrinia nubilalis*). Genetics 176:2343–2355

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35

McLeod DGR (1978) Genetics of diapause induction and termination in the European corn borer, *Ostrinia nubilalis* (Lepidoptera: pyralidae), in Southwestern Ontario. Can Ent 110:1351–1353

Mutuura A, Monroe E (1970) Taxonomy and distribution of the European corn borer and allied species: genus *Ostrinia* (Lepidoptera: Pyralidae). Mem Entomol Soc Can 71:1–112

Przeworski M (2002) The signature of positive selection at randomly chosen loci. Genetics 160:1179–1189

Przeworski M, Wall JD, Andolfatto P (2001) Recombination and the frequency spectrum in *Drosophila melanogaster* and *Drosophila simulans*. Mol Bio Evol 18:291–298

Reed GL, Guthrie WD, Showers WB, Barry BD, Cox DF (1981) Sex-linked inheritance of diapause in the European corn borer: its significance to diapause physiology and environmental response of the insect. Ann Entomol Soc Am 74:1–8

Roelofs WL, Du JW, Tang XH, Robbins P, Eckenrode C (1985) Three European corn borer populations in New York based on sex pheromones and voltinism. J Chem Ecol 11:829–836

Roelofs WL, Glover T, Tang XH et al (1987) Sex pheromone production and perception in European corn borer moths is determined by both autosomal and sex-linked genes. Proc Natl Acad Sci USA 84:7585–7589

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics 160:753–763

Schluter D (2000) The ecology of adaptive radiation. Oxford University Press, Oxford

Showers WB, Chiang HC, Keaster AJ, Hill RE, Reed GL, Sparks AN, Musick GJ (1975) Ecotypes of the European corn borer in North America. Environ Entomol 4:753–760

Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. Genet Res 71:155–160

Smith HE (1920) Broom corn, the probable host in which *Pyrausta nubilalis* Hubn. reached America. J Econ Entomol 13:425–430

Stajich JE, Hahn MW (2005) Disentangling the effects of demography and selection in human history. Mol Biol Evol 22:63–73

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–170

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol 14:671–688

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269–285

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thomas Y, Bethenod M, Pelozuelo L, Frérot B, Bourguet D (2003) Genetic isolation between two sympatric host-plant races of the European corn borer, *Ostrinia nubilalis* Hübner. I. Sex pheromone, moth emergence timing, and parasitism. Evolution 57:261–273

Thornton K (2003) libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19:2325–2327

Thornton K, Andolfatto P (2006) Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics 172:1607–1619

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