



# Whole-genome data reveal the complex history of a diverse ecological community

Lynsey Bunnefeld<sup>a,b,1,2</sup>, Jack Hearn<sup>a,1</sup>, Graham N. Stone<sup>a,3</sup>, and Konrad Lohse<sup>a,3</sup>

<sup>a</sup>Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, Scotland; and <sup>b</sup>Biological & Environmental Sciences, University of Stirling, Stirling FK9 4LA, Scotland

Edited by Craig Moritz, Australian National University, Canberra, ACT, Australia, and accepted by Editorial Board Member Douglas Futuyma May 30, 2018 (received for review January 8, 2018)

How widespread ecological communities assemble remains a key question in ecology. Trophic interactions between widespread species may reflect a shared population history or ecological fitting of local pools of species with very different population histories. Which scenario applies is central to the stability of trophic associations and the potential for coevolution between species. Here we show how alternative community assembly hypotheses can be discriminated using whole-genome data for component species and provide a likelihood framework that overcomes current limitations in formal comparison of multispecies histories. We illustrate our approach by inferring the assembly history of a Western Palearctic community of insect herbivores and parasitoid natural enemies, trophic groups that together comprise 50% of terrestrial species. We reject models of codispersal from a shared origin and of delayed enemy pursuit of their herbivore hosts, arguing against herbivore attainment of “enemy-free space.” The community-wide distribution of species expansion times is also incompatible with a random, neutral model of assembly. Instead, we reveal a complex assembly history of single- and multispecies range expansions through the Pleistocene from different directions and over a range of timescales. Our results suggest substantial turnover in species associations and argue against tight coevolution in this system. The approach we illustrate is widely applicable to natural communities of nonmodel species and makes it possible to reveal the historical backdrop against which natural selection acts.

population genomics | comparative phylogeography | gall wasps | chalcid parasitoids | Western Palearctic

**E**cological communities comprise organisms of multiple species living in a specified place and time (1). A common feature of life on Earth is that very similar communities, showing high overlap in species composition and patterns of interaction, can be found in multiple locations (2). How such sets of communities assemble remains largely unknown (3). Are similar communities derived by expansion of a single original community, with component species expanding their distributions side by side such that ecological interactions among them are maintained? Or do similar communities arise repeatedly through a process of ecological fitting (2, 4), involving the local sorting of species pools into communities based on trait values that have not been shaped by coevolution? Alternatively, community assembly could be an entirely random and ecologically neutral process (5). Which of these models applies is central to our understanding of how traits associated with species interactions evolve, because while codispersal allows sustained coadaptation and coevolutionary interactions among component species, no such processes are associated with ecological fitting or random assembly. These alternative models also make contrasting predictions for variation in population history across component species in a community. While codispersal predicts a concordant origin and direction of range expansion in interacting species, ecological fitting and neutral processes have no such requirement and can accommodate very discordant population histories.

Concordant population histories are expected in obligate species associations, such as those between plants and specialist pollinators, and these commonly show both coevolution of associated traits and codiversification of the lineages involved (6, 7). However, much of terrestrial diversity is found in communities dominated by less specific interactions between guilds of species, for which either codispersal or ecological fitting is a plausible assembly mechanism. These are exemplified by the rich insect communities associated with temperate trees, in which widespread herbivores are commonly attacked by a consistent set of parasitoid enemies (8, 9). Both herbivores and parasitoids in such systems show traits that structure trophic interactions (10–12), but the extent to which these represent coadaptations or the results of ecological fitting remains little understood. Indeed, evidence exists for asynchronous divergence between Joshua tree and yucca moth populations, suggesting that, despite the obligate nature of this pollination mutualism, delayed host tracking may also be at play (13).

Here we assess the evidence for four alternative models (Fig. 1 *A–D*) in the assembly of an exemplar community of oak-associated insects, comprising herbivorous insect hosts and

## Significance

Widespread biological communities are common, but little is known about how they assemble. A key question is how sets of trophically linked species (predators and their prey, hosts and parasites) spread to occupy current distributions. Do they disperse together, preserving ecological interactions, or separately, such that interactions are interrupted? This is central to assessing the potential for coevolution in a system and requires inference of species associations both over space and through time. Here, we use *de novo* genomic data and likelihood-based approaches to infer the assembly history of a multispecies community of Western Palearctic insect herbivores and parasitoid natural enemies—the two trophic groups that together comprise 50% of all animal species.

Author contributions: G.N.S., and K.L. designed research; L.B., J.H., G.N.S., and K.L. performed research; J.H. and K.L. contributed new reagents/analytic tools; L.B. and K.L. analyzed data; and L.B., G.N.S., and K.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. C.M. is a guest editor invited by the Editorial Board.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: Raw reads have been deposited in the European Nucleotide Archive (ENA) (accession no. [ERP023079](https://www.ebi.ac.uk/ena/record/ERP023079)) and the Short Read Archive (accession no. [PRJEB20883](https://www.ncbi.nlm.nih.gov/sra/PRJEB20883)). Genome assemblies are deposited in the ENA (accession nos. [PRJEB27189](https://www.ebi.ac.uk/ena/record/PRJEB27189) and [ERP109243](https://www.ebi.ac.uk/ena/record/ERP109243)). Python and Mathematica code can be found at <https://github.com/KLohse/BunnefeldEtAL.2018>.

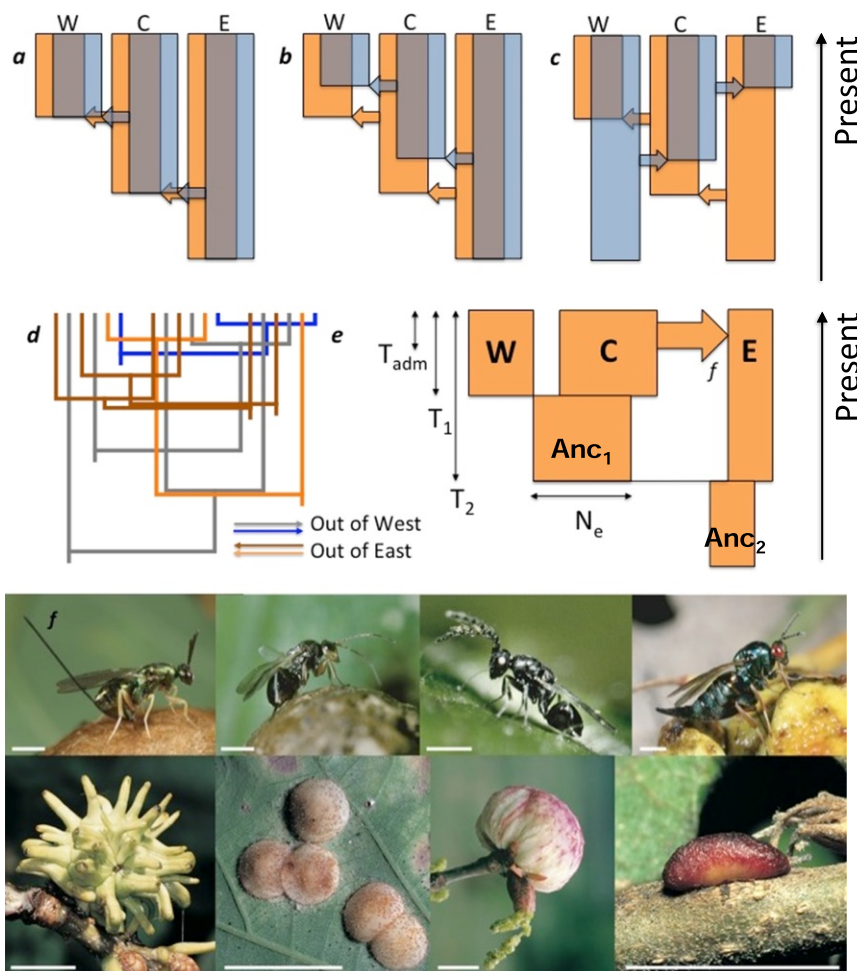
<sup>1</sup>L.B. and J.H. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: [lynsey.bunnefeld@stir.ac.uk](mailto:lynsey.bunnefeld@stir.ac.uk).

<sup>3</sup>G.N.S. and K.L. contributed equally to this work.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800334115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800334115/-DCSupplemental).

Published online June 26, 2018.



**Fig. 1.** Population histories under alternative hypotheses of community assembly. (A) Under strict codispersal species disperse between refugia together from a shared origin and their population histories show the same topology and timing of population splits. (B) Host tracking predicts that hosts and parasitoids share a common geographic origin, but allows parasitoids to follow after their hosts. Host tracking is significant ecologically because hosts can achieve a measure of enemy-free space (20) which decouples coevolutionary interactions between trophic levels. Enemy escape has been seen over ecological timescales (21, 22), but has rarely been studied in the context of population histories (20). (C) Ecological fitting results in the same set of interacting species in each refuge, but allows highly discordant patterns of range expansion from different origins. (D) Under an ecological and biogeographic null model, expansion times are random draws from a single community-wide distribution. Shown is a six-species community with two parasitoids (orange) and one gall wasp host (brown) expanding out of the east and two parasitoids (gray) and one gall wasp (blue) expanding out of the west. Split times ( $T_1$  and  $T_2$ ) are drawn randomly from an exponential distribution. (E) The demographic history of each species is captured by a seven-parameter model. We assume a separate population size ( $N_e$ ) for each refugial population (three parameters). In this example,  $N_E < N_W < N_C$  (shown by width of bars and where  $N_E$  is the  $N_e$  for the eastern refugium, and so on) and an instantaneous admixture event (shown by a horizontal arrow) at  $T_{adm}$  transfers a fraction  $f$  of the central population into the east. The ancestral populations always share an  $N_e$  with one of their descendant populations. Here,  $Anc_1$  shares an  $N_e$  with the central population and  $Anc_2$  shares an  $N_e$  with the eastern population. (F) Exemplar members of the oak gall community. Galls induced by the four oak gall wasp species (Bottom) are attacked by a range of parasitoid wasps (Top). Natural enemies from Top Left to Top Right: *Torymus auratus*, *Synergus umbraculus*, *Eurytoma brunniventris*, and *Ormyrus nitidulus*. (*S. umbraculus* is an inquiline oak gall wasp that inhabits galls induced by other oak gall wasps; for brevity we group it with parasitoid species hereafter). Galls from Bottom Left to Bottom Right: *Andricus grossulariae*, *Neuroterus quercusbaccarum*, *Biorhiza pallida*, and *Pseudoneuroterus saliens*. [Scale bars: 1 mm (Top row) and 1 cm (Bottom row).]

parasitoid and inquiline natural enemies (Fig. 1 legend). These models include strict codispersal (simultaneous range expansion), host tracking (parasitoids' pursuit of their hosts), and ecological fitting (local recruitment of parasitoids from alternative hosts), each of which makes contrasting predictions for spatial patterns of genetic variation within and among species. We also evaluate the support for an alternative null model of assembly under ecological neutrality in which a species' history is a random draw from a community-wide distribution regardless of its trophic interactions with other community members (5, 14).

Our exemplar community comprises oak cynipid gall wasp herbivores and their chalcid parasitoid wasp natural enemies, which form a set of interacting species whose distributions span the

Western Palearctic from Iberia to Iran. Associated parasitoids attack only cynipid galls, making this system ecologically closed and hence suitable for analysis in isolation (8). Single-species analyses of community members (15–18) have found westward declines in genetic diversity that support expansion into Europe from Western Asia, an “out of the East” pattern that is concordant with inferred divergence of Western Palearctic gall wasp lineages in Western Asia 5–7 Mya (19). However, the extent to which this is true of the whole community, and particularly the parasitoids, remains unclear (20).

We test the predictions of each model of community assembly using whole-genome sequence (WGS) data for 13 species (4 gall wasp herbivores and 9 parasitoids) of the oak gall community,





**Table 1. Maximum composite-likelihood estimates (MCLE) of demographic parameters and population topologies under the best supported model for each species**

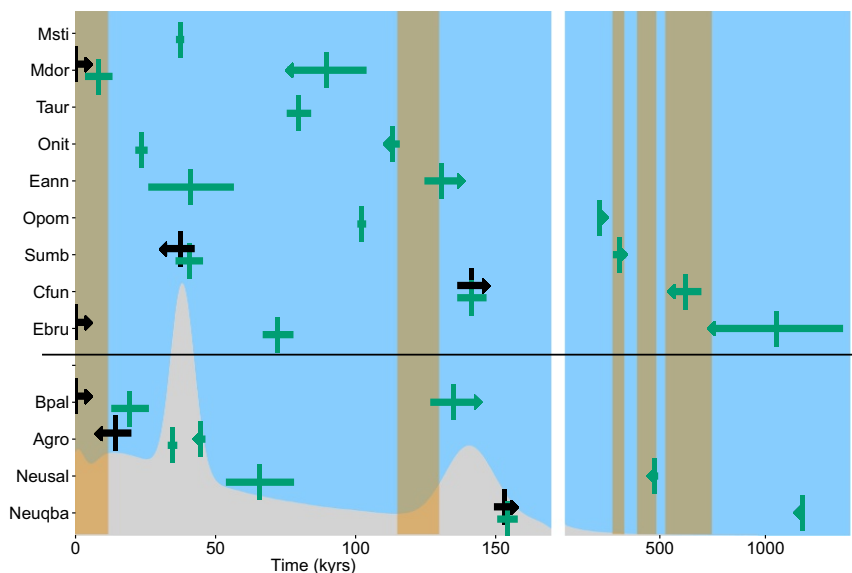
| Species binomial                  | Species code | $Anc_1$ | $Anc_2$ | Topology | $N_W$ | $N_C$ | $N_E$ | $T_{adm}$ | $T_1$ | $T_2$ | $f$  | $So \rightarrow Si$ |
|-----------------------------------|--------------|---------|---------|----------|-------|-------|-------|-----------|-------|-------|------|---------------------|
| <b>Parasitoids</b>                |              |         |         |          |       |       |       |           |       |       |      |                     |
| <i>Cecidostiba fungosa</i>        | Cfun         | West    | West    | East     | 22.9  | 367.5 | 25.0  | 142**     | 142   | 622   | 0.79 | $W \rightarrow E$   |
| <i>Eupelmus annulatus</i>         | Eann         | East    | West    | West     | 15.5  | 82.1  | 40.4  | 0         | 41    | 131   | 0.03 | —                   |
| <i>Eurytoma brunniventris</i>     | Ebru         | Centre  | Centre  | East     | 1162  | 47.5  | 116.3 | 0*        | 72    | 1051  | 0.91 | $W \rightarrow E$   |
| <i>Megastigmus dorsalis</i>       | Mdor         | Centre  | East    | East     | 2.5   | 34.3  | 11.4  | 0*        | 8     | 90    | 0.39 | $C \rightarrow E$   |
| <i>Megastigmus stigmatizans</i>   | Msti         | —       | West    | Polytomy | 6.4   | 1.0   | 0.4   | —         | 37    | 37    | —    | —                   |
| <i>Ormyrus nitidulus</i>          | Onit         | Centre  | Centre  | East     | 3.6   | 17.9  | 10.6  | 24        | 24    | 113   | 0    | —                   |
| <i>Ormyrus pomaceus</i>           | Opom         | East    | East    | West     | 18.0  | 71.3  | 30.3  | 53        | 102   | 217   | 0.06 | —                   |
| <i>Synergus umbraculus</i>        | Sumb         | Centre  | West    | West     | 8.7   | 13.4  | 19.8  | 38**      | 41    | 309   | 0.54 | $C \rightarrow W$   |
| <i>Torymus auratus</i>            | Taur         | —       | East    | Polytomy | 1050  | 20.9  | 41.7  | —         | 80    | 80    | —    | —                   |
| <b>Gall wasp hosts</b>            |              |         |         |          |       |       |       |           |       |       |      |                     |
| <i>Andricus grossulariae</i>      | Agro         | Centre  | East    | East     | 1.0   | 1.9   | 13.9  | 14        | 35    | 45    | 0.09 | $E \rightarrow W$   |
| <i>Biorhiza pallida</i>           | Bpal         | East    | West    | West     | 13.0  | 20.0  | 65.0  | 0*        | 19    | 135   | 0.12 | $W \rightarrow E$   |
| <i>Neuroterus quercusbaccarum</i> | Neuqba       | Centre  | East    | East     | 10.3  | 20.7  | 12.2  | 153**     | 154   | 1175  | 0.34 | $C \rightarrow E$   |
| <i>Pseudoneuroterus saliens</i>   | Neusal       | Centre  | Centre  | East     | 2.4   | 35.7  | 11.0  | 60        | 66    | 474   | 0.05 | —                   |

Divergence ( $T_1$ ,  $T_2$ ) and admixture times ( $T_{adm}$ ) are given in thousands of years (ky), and effective population sizes ( $N_W$ ,  $N_C$ ,  $N_E$ ) are  $\times 10^4$  individuals.  $Anc_1$  and  $Anc_2$  indicate which current population size is shared with the younger and older ancestral populations, respectively (Fig. 1E). For  $T_{adm}$ , \* indicates that the 95% confidence interval (CI) of the estimate for this parameter includes zero, while \*\* indicates that the 95% CI overlaps the 95% CI for  $T_1$ .  $So \rightarrow Si$  indicates the source and sink population for the admixture event, where the 95% CI for  $f$  was found to not overlap zero.

*M. stigmatizans* (Table 1). Given that a large number of species have nonoverlapping confidence intervals (computed using a full parametric bootstrap; *Materials and Methods*) for this split, we can confidently reject assembly via strict codispersal (Fig. 2). Irrespective of whether we consider all species jointly or only those with a putative eastern origin, the deepest population split ( $T_2$ ) is not consistently older in gall wasp hosts than in their parasitoid enemies (Fig. 2), arguing against both delayed host tracking and widespread enemy escape. We stress that our comparison of population divergence times across species does not rely on any absolute calibration but assumes only (i) knowl-

edge of the generation times of each species (which we have; *Materials and Methods* and ref. 20) and (ii) an equal mutation rate across species (which seems reasonable) (see *SI Appendix, Table S3* for uncalibrated parameter estimates for the top three best-supported models for each species).

**Testing Ecologically Neutral Assembly.** Our rejection of the codispersal and delayed host tracking models raises the question of whether the observed variation in species' demographic histories is nonetheless structured or whether it is compatible with a neutral model of community assembly. We address the first



**Fig. 2.** Community-wide patterns of range expansion across the Western Palearctic. Estimates for nine parasitoids and four gall wasp host species are shown at *Top* and *Bottom*, respectively (for species abbreviations see Table 1). Divergence times are shown in green and comprise a point estimate flanked by 95% CIs. The arrow on the older divergence time for each species shows the direction of range expansion; e.g., a right-pointing arrow indicates expansion from west to east. Note that the arrow is absent for Msti and Taur because their histories are unresolved polytomies and hence nondirectional. Where supported in the best model, admixture times are shown in black. The arrow gives the direction of admixture, again comprising a point estimate flanked by 95% CIs. CIs are arbitrary for the three species with admixture estimated at the present day. Orange vertical bars indicate interglacials. The best-fitting community-wide mixture distribution (*Materials and Methods*) of divergence and admixture times is shown in gray. Note that the x axis on the right-hand side is compressed relative to that on the left. Point estimates and 95% CIs are tabulated in *SI Appendix, Table S6*.

question by asking how many divergence events are required to explain the demographic history of this community. Considering only the oldest split in each species ( $T_2$ ), we explored models of codivergence (using a discretized grid of the marginal support for  $T_2$ ; *Materials and Methods*). The most parsimonious model of codivergence contained three clusters of species with indistinguishably similar divergence times (*T. auratus* and *Megastigmus dorsalis* at 82 kya, *Eupelmus annulatus* and *B. pallida* at 133 kya, and *E. brunniventris* and *N. quercusbaccarum* at 1,175 kya) and seven species that each diverged independently of all others ( $\Delta \ln L = 1.8$ , 3 df,  $p = 0.308$ ; *SI Appendix, Table S4*).

A second, and we argue more meaningful, question is whether the assembly of the oak gall community is compatible with an altogether random process. In other words, can we reject an ecologically neutral null model that does not consider trophic interactions? To test this, we fitted continuous distributions to the community-wide set of divergence ( $T_1$  and  $T_2$ ) and admixture ( $T_{adm}$ ) estimates. The inference of community-wide parameters was implemented in a likelihood framework with two key features: First, we propagated the uncertainty in species-specific parameter estimates by using a discretized grid of marginal support for the  $T$  parameters and  $f$  of each species. Second, for each species with significant admixture, the contributions of the time of admixture ( $T_{adm}$ ) and the oldest split ( $T_2$ ) to the community-wide distribution were weighted by its admixture fraction (i.e.,  $f$  and  $1 - f$ , respectively) (see *Materials and Methods* for details).

Arguably (14), the simplest ecologically neutral model would view the timing of divergence and admixture between refugia as a simple waiting process ( $T$  parameters are drawn from an exponential distribution) or assume a single unimodal distribution (e.g., a log normal). We were able to reject both of these null models of community assembly in favor of a more complex history involving several continuous multispecies divergence pulses. The most parsimonious model assumes that, community-wide, the times of between-refuge population divergence and admixture are drawn from a mixture distribution consisting of a single exponential and three log-normal distributions (Fig. 2) with modes at 13 kya, 38 kya, 140 kya [ $\Delta \ln L$  (relative to a model with a single exponential) = 16.7, 9 df,  $P = 0.05$ ; *SI Appendix, Table S5*].

**Community-Wide Expansion Pulses Coincide with Warm Periods in the Pleistocene.** The community-wide modes of divergence and admixture times coincide approximately with warm periods in the late Pleistocene (corresponding to the beginning of the current Holocene interglacial, a Dansgaard-Oeschger interstadial event, and just before the Eemian interglacial). This suggests that, within the oak gall community, expansions into and admixture between refugia are most likely associated with periods of geographic expansion of the distributions of oaks and their associated gall communities—a process confirmed for the Eemian interglacial by fossils for oaks, gall wasps, and parasitoids (30). Although the correspondence of community-wide time estimates with Pleistocene climatic events is striking, we stress that these absolute time estimates rely on a calibration with a *Drosophila* mutation rate and should be interpreted with caution.

## Discussion

Understanding how complex natural communities have assembled requires a retrospective approach and inference of the population histories of component species. Previous comparative analyses of the phylogeographic history of trophically linked species either have been limited in scope by the small numbers of species compared (e.g., refs. 17 and 31) and/or were based on small numbers of loci and achieved low resolution of the demographic histories of individual species (20). Here we use whole-genome data for small samples of individuals to make detailed, continental-scale reconstructions of the Pleistocene

histories of 13 members of a widespread insect community: oak gall wasps and their associated parasitoid enemies.

Our systematic comparison of population histories within and between trophic levels unveils considerable complexity in the assembly of the oak gall community. This can neither be explained by classic models of deterministic community assembly which predict a small number of codivergence events (32) nor be captured by an ecologically neutral process which predicts a simple community-wide distribution of divergence times (14).

Members of the oak gall community differ in both the directionality and timing of their longitudinal expansion into Europe, without any consistent difference between gall wasp hosts and their parasitoids. We can therefore confidently reject both codivergence and enemy escape, two contrasting and paradigmatic models of community assembly. Moreover, the fact that we identify a minimum of 10 distinct divergence (or codivergence) events for the oldest split implies that we can also rule out more complex assembly scenarios that involve a small number of codivergence events, e.g., a history where all species codiverge with at least one other. Our comparative analysis also allows us to reject an ecologically neutral null model of random community assembly. This model views species as exchangeable and assumes that divergence and admixture events between refugia happen at a single community-wide rate (5, 14) and independently of past climatic events.

**Ecological Fitting and Host Range.** Our finding that species differ in both the timing and the direction of range expansions supports the existence of a continental-scale geographic mosaic of species interactions (27, 28) whose structure has varied through time. For example, we infer that the seven species showing an out-of-the-east population history have shared the eastern refuge throughout their history. In contrast, interactions in the center and the west between any two of the same set of species are possible only when both have reached these refugia. Because divergence times in both guilds vary by an order of magnitude, there must have been considerable turnover in species interactions. For example, *E. brunniventris*, which has the most ancient out-of-the-east parasitoid population history, can only have exploited hosts whose expansion histories into the center are at least as early (such as *N. quercusbaccarum*). Interactions in the center with later-expanding hosts, such as *A. grossulariae*, could be restored only when these, too, arrived from the east. The complex temporal dynamics of species associations in this system suggest that current values of phenotypic traits associated with trophic interactions likely reflect a complex selective history, with low potential for tight coevolution between any given pair of species. Such turnover in trophic relationships may also explain why this and other communities centered around temperate insect herbivores are dominated by relatively generalist species that attack a range of hosts within the community (11, 33). We suggest that generalists are likely to have a larger fitness space (4) over host diversity than specialists, facilitating recruitment to regionally varying host assemblages through ecological fitting.

An alternative explanation for the incongruence in timing we observe across species is that the genetic signatures of older demographic events have been erased by extensive subsequent admixture in some species, but not in others. While such potential loss of signal is a general limitation of demographic inference and cannot be ruled out, it seems an unlikely explanation for our general finding of incongruence between species, given the common absence of admixture signals.

A species' potential to expand its range is a function of its ecology and life history. A recent study identified life-history strategy as the only correlate of genetic diversity: Animals with more rapid life histories and larger numbers of offspring had higher genetic diversity (25). While this and other comparative

studies of genetic diversity (26, 34) have ignored demography, comparing species with respect to their demographic histories makes it possible (at least in principle) to test for assembly rules. For example, if ecological fitting plays an important role in community assembly, more generalist and/or widespread parasitoids should have older histories than specialists. We do indeed find that  $T_2$  and ancestral  $N_e$  are positively correlated with both host range and abundance, which is compatible with this idea (*SI Appendix*, Fig. S2). However, confirming this relationship will require larger samples of species.

**Eastern Origins and Westerly Winds.** A majority of members of the oak gall community have an eastern origin, a pattern that is compatible with the east to west declines in genetic diversity found by previous studies of this community (16, 17, 31) and Western Palearctic taxa more generally (35, 36). It also matches the inferred origin of gall wasp lineages now found across the Western Palearctic ca. 10 Mya in Anatolia/Iran (19). The longitudinal colonization history of deciduous oaks in the Western Palearctic is currently too poorly resolved to know whether gall wasps tracked their oak hosts westward into Europe (37, 38). While only a few species in the oak gall community show evidence for admixture, it is intriguing that our inference of admixture is (i) most substantial in the three species with the oldest histories, *N. quercusbaccarum*, *Cecidostiba fungosa*, and *E. brunniventris* (Table 1), and (ii) directed predominantly from the west into the east, that is, in the opposite direction of population divergence. We conducted a simple simulation study to confirm that this directional signal is not an artifact of our modeling approach; i.e., admixture back into the source population is not easier to detect than in the reverse direction (*Materials and Methods*). Inference of gene flow between refugia is compatible with the documented range expansion by both gall wasps and associated parasitoids over similar spatial scales (17, 39) and the ability of gall wasps to reach offshore islands (40). Given that both chalcid parasitoids and cynipids have been observed in aerial samples taken 200 m from the ground (41) and the prevailing winds in the Western Palearctic are from the west, it seems plausible that gene flow into the east is simply easier and therefore more likely than in the reverse direction.

**Limitations and Potential Confounding Factors.** While the demographic models we consider are necessarily simplifications of the truth, they capture key population-level processes that have been largely ignored by previous comparative analyses. First, many comparative phylogeographic studies have focused on divergence between only pairs of populations (20, 42). Second, analyses of divergence and admixture between multiple populations have, for the sake of tractability, often assumed equal  $N_e$  among populations (43, 44) which can bias estimates of divergence and admixture. Our finding that  $N_e$  varies by up to two orders of magnitude among populations within species (Table 1) highlights the importance of modeling  $N_e$  parameters explicitly (45, 46). While this study's reanalysis of *B. pallida* inferred little admixture when accounting for  $N_e$  differences between refugia, a previous analysis based on a single sample per population (43) suggested substantial gene flow from east to west under a model in which ancestral  $N_e$  was fixed across all populations.

To make robust inferences about community assembly, it is imperative to minimize biases that might affect comparability among species. We have sequenced species to the same coverage whenever possible and have used the same bioinformatic pipeline and comparable block lengths (in terms of diversity) for all species. To test whether selective constraint is likely to differ drastically among species or trophic levels (which could lead to systematic biases in demographic estimates), we compared diversity across all sites to that at synonymous coding sites and found no difference between hosts and parasitoids (*SI Appendix*, Suppl.

Info 2). Finally, our parametric bootstrap replicates which were simulated with recombination show that recombination (which may differ between species) had minimal effects on model and parameter estimates.

**Implications for Comparative Phylogeography.** Previous comparative studies have been unable to build such a nuanced picture of community assembly in this or any other system due to limitations in both available data and analytical approaches (20, 47, 48) [although new methods have begun to emerge (49, 50); see ref. 13 for an early successful example]. In particular, we conducted a previous analysis of the oak gall wasp community based solely on mitochondrial data sampled for the same set of refugia and an overlapping set of taxa. In contrast to the present study, this previous analysis had very limited power to resolve even simple demographic histories with or without gene flow (20).

We now have both the inference tools and the data necessary to resolve divergence and admixture relationships among multiple populations and across many trophically linked species. Given the resolution of WGS data, it becomes hard to justify the logic of testing for strict codivergence (51), if this implies identical divergence times across species. Three recent comparative phylogeographic analyses, each based on thousands of ddRAD loci, avoided using approaches that test for strict codivergence, but instead compared parameter estimates post hoc among species (52–54). For example, Oswald et al. (53) found that six pairs of codistributed birds shared the same mode of divergence (isolation with asymmetric migration) but diverged asynchronously. Our approach replaces informal post hoc comparisons with a formal comparative framework that both incorporates uncertainty in species-specific parameter estimates and characterizes community-wide distributions of divergence times.

**Outlook.** Being able to efficiently reconstruct intraspecific histories from genome-scale data across multiple species means that one can infer the tempo and mode of community assembly across space and through time. The composite-likelihood framework we use applies to a wide range of demographic histories (31, 55) and, as we have shown here, lends itself to comparative analyses. Given sufficient replication in terms of species, such analyses open the door to answering a range of fundamental questions in evolution and ecology. For example, an interesting avenue of future research will be to explicitly include selection in comparative analyses. In particular, it will be fascinating to ask whether and how range shifts and admixture events are accompanied by selection on genes that may be involved in host–parasitoid interactions. It will soon be possible to use WGS data to compare not only the demographic histories of interacting species but also the signatures of selective sweeps and their targets in the genome. In practice, such more detailed comparisons will require both further work on inference methods and more complete datasets, including more contiguous (and functionally annotated) reference genomes and larger samples of individuals and taxa.

## Materials and Methods

### Samples and Sequencing.

**Sample processing.** We generated WGS data for three species of oak gall wasp hosts, eight parasitoids, and one inquiline. We also reanalyzed WGS data generated for a pilot study for *B. pallida*, another gall wasp species (43) (ENA Sequence Read Archive: <http://www.ebi.ac.uk/ena/data/view/ERP002280>). For each species we sampled a minimum of two male individuals from each of three refugial populations spanning the Western Palearctic: Iberia (west), the Balkans (center), and Iran (east) (*SI Appendix*, Table S2). Only a single eastern sample was available for two species, *Ormyrus pomaceus* and *B. pallida*, and a single western sample for *S. umbraculus*.

DNA was extracted from (haploid) male individuals, using the Qiagen DNeasy kit. Nextera libraries were generated for each individual specimen



and sequenced on a HiSeq 2000 by Edinburgh Genomics. Raw reads are deposited in the European Nucleotide Archive (accession nos. PRJEB20883/ERP023079). Median average coverage per individual ranged from 3.81 to 6.59 for the parasitoids and from 2.94 to 9.33 for the gall wasps. The inquiline gall wasp *S. umbraculus* was sequenced to higher coverage (median 34.1) because its genome size was initially overestimated.

**De novo genome assembly.** Reads were quality controlled and adapter trimmed using trimmomatic (56), followed by cutadapt (57) to remove remaining Nextera transposase sequences. Results were inspected using FastQC (58). For each species, data were combined across all individuals and a reference genome assembly generated using SPAdes (59), MaSuRCA (60), or CLC de novo assembler (SI Appendix, Table S1). Assemblies are available at the European Nucleotide Archive (accession nos. PRJEB27189/ERP109243). We used RepeatScout (61) and RepeatMasker (62) to de novo predict repeat regions. Contigs that Blast (63) matched putative contaminant (bacteria, fungi) or mitochondrial sequence after inspection of blobplots (64) were removed from further analysis.

**Variant calling.** Reads per individual were mapped back to each reference genome using BWA (0.7.15-r1140) (65) and duplicates marked using picard ([broadinstitute.github.io/picard/](http://broadinstitute.github.io/picard/)) (V2.9.0) MarkDuplicates. Variant calling was performed with freebayes (<https://github.com/ekg/freebayes>) (v1.1.0-3-g961e5f3) with minimum alternate allele count (-C) of 1, minimum base quality (-q) of 10, and minimum mapping quality (-q) of 20. Variants were then filtered for a minimum quality of 20 and decomposed into allelic primitives. Remaining single-nucleotide polymorphisms (SNPs) were retained for further analyses.

**Generating blockwise data.** The CallableLoci walker of GATK (v3.4) (66) was used to identify regions covered in all individuals and with a base quality >10 and a mapping quality >20 in the mark-duplicated bam files (these filters are the same used for SNP calling in freebayes). Blocks of a fixed length *l* of callable sites were generated for each species, using custom scripts. For parasitoids, we included only sites that passed the CallableLoci filter in all six individuals. Given the overall lower sequencing depth for the gall wasps, we relaxed this requirement in one of the four gall wasp hosts, *Pseudoneuroterus saliens*, and partitioned the data into blocks using the CallableLoci filter independently for each subsample of *n* = 3 (see below).

Since our inference framework assumes no recombination within blocks, we partitioned the data into short blocks (and performed a simulation-based check for potential biases). To ensure equal information content and to minimize differences in bias due to recombination within blocks across species, we generated blocks with a fixed average number of 1.5 pairwise differences. In other words, the physical length of blocks *l* was chosen to be inversely proportional to the average pairwise per site diversity  $\pi$  (postfiltering) of a species and so differed between taxa.

Blocks with a physical span (prefiltering length) of >2*l* or containing more than five “None” sites were removed. Likewise, short contigs (<2*l*) were ignored. These filtering steps resulted in datasets of between 56,031 and 659,990 blocks per species. For *P. saliens* where each subsample was filtered independently, average block number was 78,105 blocks.

### Fitting Demographic Models.

**Computing model likelihoods.** Models were fitted to block-wise data using an analytic coalescent framework (24). Briefly, the probability of observing a particular mutational configuration in a sequence block can be expressed as a higher-order derivative of the generating function (GF) of genealogical branch lengths. This calculation assumes an infinite-sites mutation model (which is reasonable given the low per site diversity). It is straightforward to calculate the support (i.e., the logarithm of the likelihood, *lnL*) of a particular model, given the corresponding GF and a table of counts of observed block-wise mutational configurations.

We used an automation (23) previously implemented in Mathematica (67) to obtain the GF for the set of 48 models. Since a solution for the GF for the full sample of six individuals (two per population) is intractable, we computed the composite likelihood,

$$CL = \prod_i^{\# \text{blocks}} \prod_{j \subset n: j=|3|} L(\Theta | k_{ij}),$$

where  $k_{ij}$  is the count of the mutations defined by the three unrooted branches for block *i* and triplet *j*. In other words, the composite likelihood (CL) is a product over contributions from all (14 for a sample of size *n* = 6 we consider) possible (unrooted) subsamples of triplets and all blocks. This calculation scales to arbitrarily large samples of individuals as

the GF expressions for each triplet are small. It also makes use of the phase information contained in the haploid data. Unlike previous implementations (i.e., ref. 43) there is no requirement for an outgroup. Analogous CL strategies based on subsamples have been used in phylogenetic network analysis (68) and to infer divergence and continuous gene flow (69, 70). We maximized *lnCL* for all parameters ( $N_e, N_c, N_w, T_1, T_2, T_{adm}, f$ ) included in the best-supported model, using a Nelder–Mead simplex algorithm implemented in the Mathematica function NMaximize. Configurations involving more than  $k_{max} = 2$  mutations on any particular branch were combined in the CL calculation. The method differs from other recent approaches for inferring reticulate evolutionary histories in that it allows  $N_e$  to differ between populations (cf. ref. 68) and makes use of linkage among sites [cf. TreeMix (71)].

**Parametric bootstrap.** Maximizing the CL across blocks and subsamples ignores linkage between blocks but gives an unbiased point estimate of parameters. To measure the uncertainty in demographic parameter estimates we conducted a full parametric bootstrap: For each species, we simulated 100 replicate datasets in msprime (72) under the full ancestral recombination graph and the best-supported model (and the MCLs under that model). Recombination rates per base pair ( $\rho$ ) were inferred for each species, using the two-locus GF outlined in ref. 24. This inference was based on two Spanish individuals from each species. Each bootstrap replicate had the same total sequence length as the real dataset (after filtering). We assumed that blocks in the bootstrap replicates were immediately adjacent to one another, which is conservative given the filtering pipeline used to delimit blocks in the real data. For the sake of computational efficiency, each simulation replicate was divided into 50 equally sized chunks/chromosomes. The 95% CIs were obtained as 2 SDs of estimates across bootstrap replicates.

**Identifying codivergence clusters.** To assess whether the oldest divergence times  $T_2$  differed significantly between species and trophic levels, we implemented a hierarchical clustering procedure. For each of the six species that overlapped in a 95% CI for  $T_2$  with any other species, we evaluated the marginal *lnCL* along a discretized grid (maximizing *lnCL* across all other parameters). We used the SD in  $T_2$  estimates across parametric bootstrap replicates to rescale these marginal *lnCL*. These rescaled *lnCL* were treated as an approximation of the true marginal support (*lnL*) for  $T_2$ .

To test whether divergence times differed between species, we computed the support for models in which divergence times are shared by subsets of taxa. Essentially, this phrases the test for clustered divergence times as a model simplification problem: For example, the MCL for a codivergence event involving two or more species is given by the maximum of the sum of their respective marginal *lnL*. The associated reduction in model support is measured by  $\Delta \ln L$  relative to the full model, i.e., allowing species to have unique divergence times. We identified the most parsimonious model as the model with the fewest distinct  $T_2$  clusters (assuming  $\Delta \ln L$  has a  $\chi^2$  distribution) and examined all possible combinations of codiverged species.

**Characterizing the community-wide distribution of divergence and admixture times.** We fitted a series of distributions of increasing complexity to the inferred *T* parameters for all species. We assumed that divergence and admixture times across species are given by some community-wide function  $g(t; \lambda)$ . The likelihood of  $\lambda$ , the community-wide rate of population divergence (or set of parameters in the case of more complex functions) is

$$L(\lambda) = \prod_{k=1}^n \int_{f_k} \int_{T_{adm,k}} \int_{T_{1,k}} \int_{T_{2,k}} p(\Theta_k) f_k g(T_{adm,k}; \lambda) \times g(T_{1,k}; \lambda) \times (1 - f_k) g(T_{2,k}; \lambda) \quad [1]$$

where  $\Theta_k$  is the set of time parameters and the admixture fraction for the *k*th species and  $p(\Theta_k)$  is the (normalized) likelihood of  $\Theta_k$ . Since computing Eq. 1 directly from the data is intractable, we used a discretized grid of likely *T* and *f* values (i.e., replacing the integrals in [1] by sums). Grids were centered around the MCL and bounded by the 95% CI of parameters. To ensure that each species contributes equal information to the community-wide inference, we normalized the *lnCL* grid of each species assuming a total of 2,000 unlinked blocks. Although admittedly arbitrary, this is very conservative: If we assume that blocks are randomly distributed in the genome, this normalization corresponds to an average distance of at least 150 kb between blocks (assuming a minimum genome size of 300 Mb). In each dimension, we evaluated five points. For computational tractability, we fixed  $\theta$  and the  $N_e$  scalars at their MCLs.

We initially fitted exponential and log-normal distributions, followed by increasingly complex mixture distributions of a single exponential and several log normals. Models were fitted sequentially: As we included additional mixture contributions, we fixed the parameters and mixture

weights of the lowest weighted component of the previous distribution. We identified the most parsimonious model as the simplest model (i.e., with the fewest parameters) which did not give a significant reduction in *lnL*. The python scripts and Mathematica notebooks are available at <https://github.com/KLohse/BunnefeldETAL2018>.

**ACKNOWLEDGMENTS.** We thank Richard Harrison, James Nicholls, Sarah White, and Lisa Cooper for help in the laboratory and Edinburgh Genomics for sequencing. We give many thanks to Stuart Baird, Alex Hayward, and

Mike Hickerson for discussions throughout the project and to Ally Phillimore for a critical manuscript review. We are grateful to Rob Ness for advice on the bioinformatics; to Pablo Fuentes-Utrilla, György Csóka, George Melika, and Majide Tavakoli for contributing specimens; to Richard R. Askew for help with identification; and to William Walton for collating the ecological data. This work was supported by a standard grant from the Natural Environmental Research Council (NERC) United Kingdom (to G.N.S. and K.L.) (NE/J010499/1). K.L. is supported by a NERC Independent Research fellowship (NE/L011522/1).

- Vellend M (2010) Conceptual synthesis in community ecology. *Q Rev Biol* 85:183–206.
- Janzen DH (1985) On ecological fitting. *Oikos* 45:308–310.
- Ricklefs RE (2015) Intrinsic dynamics of the regional community. *Ecol Lett* 18:497–503.
- Agosta SJ, Klemens JA (2008) Ecological fitting by phenotypically flexible genotypes: Implications for species associations, community assembly and evolution. *Ecol Lett* 11:1123–1134.
- Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton Monographs in Population Biology (Princeton Univ Press, Princeton).
- Jousselin E, Rasplus JY, Kjellberg F (2003) Convergence and coevolution in a mutualism: Evidence from a molecular phylogeny of figs. *Evolution* 57:1255–1269.
- Mueller UG, Rehner SA, Schultz TR (1998) The evolution of agriculture in ants. *Science* 281:2034–2038.
- Stone GN, Schönrögge K, Atkinson RJ, Bellido D, Pujade-Villar J (2002) The population biology of oak gall wasps (hymenoptera: Cynipidae). *Annu Rev Entomol* 47:633–668.
- Leppänen SA, Altenhofer E, Liston AD, Nyman T (2013) Ecological versus phylogenetic determinants of trophic associations in a plant-leafminer-parasitoid food web. *Evolution* 67:1493–1502.
- Bailey R, et al. (2009) Host niches and defensive extended phenotypes structure parasitoid wasp communities. *PLoS Biol* 7:e1000179.
- Askew RR (1980) The diversity of insect communities in leaf mines and plant galls. *J Anim Ecol* 49:145–152.
- Stireman JO, Singer MS (2003) Determinants of parasitoid-host associations: Insights from a natural tachinid-lepidopteran community. *Ecology* 84:296–310.
- Smith CI, Godsoe WKW, Tank S, Yoder JB, Pellmyr O (2008) Distinguishing coevolution from covariance in an obligate pollination mutualism: Asynchronous divergence in Joshua tree and its pollinators. *Evolution* 62:2676–2687.
- Rosindell J, Hubbell SP, Etienne RS (2011) The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol Evol* 26:340–348.
- Lohse K, Sharanowski B, Stone GN (2010) Quantifying the population history of the oak gall parasitoid *Cecidostiba fungosa*. *Evolution* 58:439–442.
- Stone GN, et al. (2007) The phylogeographical clade trade: Tracing the impact of human-mediated dispersal on the colonization of northern Europe by the oak gallwasp *Andricus kollari*. *Mol Ecol* 16:2768–2781.
- Hayward A, Stone GN (2006) Comparative phylogeography across two trophic levels: The oak gall wasp *Andricus kollari* and its chalcid parasitoid *Megastigmus stigmatizans*. *Mol Ecol* 15:479–489.
- Nicholls JA, et al. (2010) Concordant phylogeography and cryptic speciation in two western Palaearctic oak gall parasitoid species complexes. *Mol Ecol* 19:592–609.
- Stone GN, et al. (2009) Extreme host plant conservatism during at least 20 million years of host plant pursuit by oak gallwasp. *Evolution* 63:854–869.
- Stone GN, et al. (2012) Reconstructing community assembly in time and space reveals enemy escape in a western palaearctic insect community. *Curr Biol* 22:531–537.
- Prior KM, Hellmann JJ (2013) Does enemy loss cause release? A biogeographical comparison of parasitoid effects on an introduced insect. *Ecology* 94:1015–1024.
- Vosteen I, Gershenzon J, Kunert G (2016) Enemy-free space promotes maintenance of host races in an aphid species. *Oecologia* 181:659–672.
- Lohse K, Chmelik M, Martin SH, Barton NH (2016) Efficient strategies for calculating blockwise likelihoods under the coalescent. *Genetics* 202:775–786.
- Lohse K, Harrison RJ, Barton NH (2011) A general method for calculating likelihoods under the coalescent process. *Genetics* 58:977–987.
- Romiguier J, et al. (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263.
- Leffler EM, et al. (2012) Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biol* 10:1–9.
- Thompson J (2005) *The Geographic Mosaic of Coevolution* (Univ of Chicago Press, Chicago).
- Hoberg EP, Brooks DR (2008) A macroevolutionary mosaic: Episodic host-switching, geographical colonization and diversification in complex host-parasite systems. *J Biogeogr* 35:1533–1550.
- Keightley PD, et al. (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res* 19:1195–1201.
- Stone GN, van der Ham RWJM, Brewer JG (2008) Fossil oak galls preserve ancient multitrophic interactions. *Proc R Soc B Biol Sci* 275:2213–2219.
- Lohse K, Barton NH, Melika N, Stone GN (2012) A likelihood-based comparison of population histories in a parasitoid guild. *Mol Ecol* 49:832–842.
- Ricklefs RE (2008) Disintegration of the ecological community. *Am Nat* 172:741–750.
- Askew RR (1961) On the biology of the inhabitants of oak galls of Cynipidae (Hymenoptera) in Britain. *Trans Soc Br Entomol* 14:237–258.
- Corbett-Detig RB, Hartl DL, Sackton TB (2015) Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol* 13:1–25.
- Koch MA, et al. (2006) Three times out of Asia minor: The phylogeography of *Arabis alpina* L. (Brassicaceae). *Mol Ecol* 15:825–839.
- Duvalx L, Belkhir K, Boulesteix M, Boursot P (2011) Isolation and gene flow: Inferring the speciation history of European house mice. *Mol Ecol* 20:5248–5264.
- Petit RJ, et al. (2002) Chloroplast DNA variation in European white oaks: Phylogeography and patterns of diversity based on data from over 2600 populations. *For Ecol Manag* 156:5–26.
- Atkinson R, Rokas A, Stone G (2007) Longitudinal patterns in species richness and genetic diversity in European oaks and oak gallwasp. *Phylogeography in Southern European Refugia*, ed Weiss S (Springer, Dordrecht, The Netherlands), pp 127–154.
- Stone GN, Sunnucks P (1993) Genetic consequences of an invasion through a patchy environment—The cynipid gallwasp *Andricus quercuscalicis* (Hymenoptera: Cynipidae). *Mol Ecol* 2:251–268.
- Walker P, Leather SR, Crawley MJ (2002) Differential rates of invasion in three related alien oak gall wasps (cynipidae: Hymenoptera). *Divers Distrib* 8:335–349.
- Chapman J, Reynolds D, Smith A, Smith E, Woivod I (2004) An aerial netting study of insects migrating at high altitude over England. *Bull Entomol Res* 94:123–136.
- Hickerson MJ, Stahl EA, Lessios HA, Crandall K (2006) Test for simultaneous divergence using approximate Bayesian computation. *Evolution* 60:2435–2453.
- Hearn J, et al. (2014) Likelihood-based inference of population history from low-coverage de novo genome assemblies. *Mol Ecol* 23:198–211.
- Yu Y, Dong J, Liu KJ, Nakhleh L (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proc Natl Acad Sci USA* 111:16448–16453.
- Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends Ecol Evol* 26:448–456.
- Hung CM, Drovetski SV, Zink RM (2017) The roles of ecology, behaviour and effective population size in the evolution of a community. *Mol Ecol* 26:3775–3784.
- Sutton TL, Riegler M, Cook JM (2016) One step ahead: A parasitoid disperses farther and forms a wider geographic population than its fig wasp host. *Mol Ecol* 25:882–894.
- Espindola A, Carstens BC, Alvarez N (2014) Comparative phylogeography of mutualists and the effect of the host on the genetic structure of its partners. *Biol J Linn Soc* 113:1021–1035.
- Beeravolu Reddy C, Hickerson MJ, Frantz LAF, Lohse K (2017) Blockwise site frequency spectra for inferring complex population histories and recombination. *bioRxiv:077958*. Preprint, posted September 25, 2017.
- Xue AT, Hickerson MJ (2015) The aggregate site frequency spectrum for comparative population genomic inference. *Mol Ecol* 24:6223–6240.
- Papadopoulou A, Knowles LL (2016) Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proc Natl Acad Sci USA* 113:8018–8024.
- Satler JD, Carstens BC (2017) Do ecological communities disperse across biogeographic barriers as a unit? *Mol Ecol* 26:3533–3545.
- Oswald JA, Overcast I, Mauck WM, Andersen MJ, Smith BT (2017) Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds. *Mol Ecol* 26:1386–1400.
- Rougemont Q, et al. (2017) Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol* 26:142–162.
- Bunnefeld L, Frantz LAF, Lohse K (2015) Inferring bottlenecks from genome-wide samples of short sequence blocks. *Genetics* 201:1157–1169.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10.
- Andrews S (2010) FastQC: A quality control application for high throughput sequencing data. Available at [www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Accessed December 12, 2013.
- Bankevich A, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.
- Zimin AV, et al. (2013) The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677.
- Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21:i351–i358.
- Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chap 4:Unit 4.10.
- Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M (2013) Blobology: Exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated gc-coverage plots. *Front Genet* 4:237.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Auwer GA, et al. (2013) From FastQ data to high confidence variant calls: The genome analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11.10.1–11.10.33.



67. Wolfram Research I (2017) *Mathematica, Version 11.1.1.0* (Wolfram Research, Inc., Champaign, IL).
68. Than C, Ruths D, Nakhleh L (2008) Phylonet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9:322.
69. Costa RJ, Wilkinson-Herbots H (2017) Inference of gene flow in the process of speciation: An efficient maximum-likelihood method for the isolation-with-initial-migration model. *Genetics* 205:1597–1618.
70. Yang Z (2010) A likelihood ratio test of speciation with gene flow using genomic data. *Genome Biol Evol* 2:200–211.
71. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:1–17.
72. Kelleher J, Etheridge AM, McVean G (2016) Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol* 12:1–22.