Networks of spatial genetic variation across species

Miguel A. Fortuna[a], Rafael G. Albaladejo[b], Laura Fernández[b], Abelardo Aparicio[b], and Jordi Bascompte[a,1]

[a]Integrative Ecology Group, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Américo Vespucio s/n, 41092 Sevilla, Spain; and [b]Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Avenida Reina Mercedes s/n, 41012 Sevilla, Spain

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved September 9, 2009 (received for review July 14, 2009)

Spatial patterns of genetic variation provide information central to many ecological, evolutionary, and conservation questions. This spatial variability has traditionally been analyzed through summary statistics between pairs of populations, therefore missing the simultaneous influence of all populations. More recently, a network approach has been advocated to overcome these limitations. This network approach has been applied to a few cases limited to a single species at a time. The question remains whether similar patterns of spatial genetic variation and similar functional roles for specific patches are obtained for different species. Here we study the networks of genetic variation of four Mediterranean woody plant species inhabiting the same habitat patches in a highly fragmented forest mosaic in Southern Spain. Three of the four species show a similar pattern of genetic variation with well-defined modules or groups of patches holding genetically similar populations. These modules can be thought of as the long-sought-after, evolutionarily significant units or management units. The importance of each patch for the cohesion of the entire network, though, is quite different across species. This variation creates a tremendous challenge for the prioritization of patches to conserve the genetic variation of multispecies assemblages.

complex networks | gene flow | habitat fragmentation | population genetics

As our influence on the biosphere keeps growing, a larger fraction of previously continuous populations become fragmented into disjunct, isolated habitat patches surrounded by a matrix of unfavorable habitat (1). Each of these patches contains a fraction of the genetic diversity of the metapopulation, and understanding the evolution and conservation of such a metapopulation hinges on understanding the spatial distribution of genetic variation (2). Without this variation, it is difficult for a population to adapt to environmental changes, which therefore makes it more prone to extinction. A critical task in the face of global change, therefore, is to map the spatial structure of this genetic variation and to relate this to its robustness to further habitat transformation.

Genetic variation is measured as the tendency of individual genotypes in a population to vary from one another. The study of the spatial structure of genetic variation is a long-standing question in population genetics (3–7). In the last few years, there has been a growing interest in understanding how geographical and environmental features structure such genetic variation, as exemplified by the new subject of landscape genetics (8, 9). More recently, this approach has benefited from a network perspective (the so-called population graphs) embracing the simultaneous statistical relationships between all populations (10). To date, those papers that have applied network theory to explain spatial patterns of genetic variation have all focused on a single species (10–14). The question now is to what extent we can generalize the conclusions of these single-species studies to other related species. From a basic point of view, it is an important question to unravel whether gene flow in space is structured similarly across species and therefore whether similar mechanisms are at work. From a more applied perspective, this is a preliminary step to assess the degree to which management strategies can be applied to multispecies assemblages or have to be applied on a species-to-species basis.

Here we analyze the spatial pattern of genetic variation in four Mediterranean shrub species in a fragmented landscape of forest patches in Southern Spain (Fig. 1). These species *(Cistus salvifolius, Myrtus communis, Pistacia lentiscus,* and *Quercus coccifera)*, have contrasting life histories and are a good representation of the woody plant species in this Mediterranean region. We focus on the 23 habitat patches inhabited simultaneously by the four plant species. We have analyzed the genetic structure of these four species by using isozymes as multivariant codominant markers (see *Materials and Methods*). Our approach is based on the integration of population graphs as a way to prune the original network of spatial genetic variation in a meaningful and informative way, and modularity analysis as a way to describe the structure of such a simplified network. This integrated approach, together with the extension to multispecies assemblages, makes our study stand out from previous papers (10–14).

From our genetic data, we start by using the method of Dyer and Nason (10) to build four networks of genetic similarity among patches, one for each plant species. The starting point is a fully connected network in which all patches are linked to each other by their genetic similarity. Dyer and Nason's method allows us to prune the original network by removing all links connecting patches whose genetic similarity is mediated by their genetic similarity with common patches (see *Material and Methods* for a step-by-step description of the statistical approach). This procedure leads to networks of genetic variation containing the smallest link set that sufficiently explains the genetic covariance structure among patches. This methodology contrasts with the pruning proposed by a recent paper based on a cutoff strength of the genetic similarity below which links were removed (14). Our method also extends Dyer and Nason's procedure by taking into account the observed allelic frequency when calculating the genetic similarity among patches. It also calculates the quantitative values of genetic similarity for the small set of links remaining in the resulting network.

Once the network of genetic similarity is constructed, we investigate its modular organization, where modules are defined genetically, not spatially. In general, a modular network is one structured in modules tightly connected internally, but loosely connected to patches from other modules (11, 15, 16). In our specific context, a module is a set of habitat patches holding populations more genetically similar to one another than to populations within patches belonging to other modules. This provides a simple description of how the genetic variation is structured in space, for each of our four species. Our ultimate goal is to assess (*i*) whether a similar modular organization is observed across the different species-specific networks and (*ii*) whether a given patch plays similar roles in these different networks of genetic variation.

Author contributions: A.A. and J.B. designed research; M.A.F., R.G.A., L.F., performed research; M.A.F., R.G.A., L.F., and A.A. contributed new reagents/analytic tools; M.A.F., R.G.A., and L.F. analyzed data; and J.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: bascompte@ebd.csic.es.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907704106/DCSupplemental.

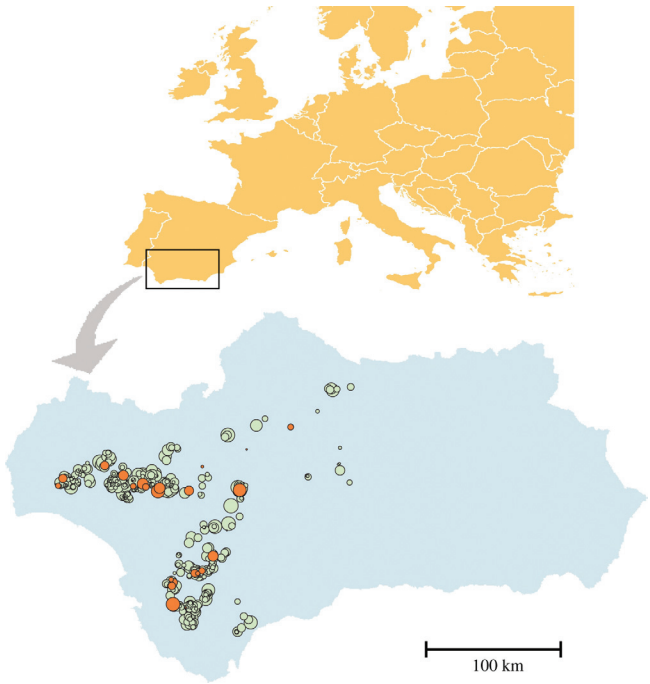

Fig. 1. Geographical location of the fragmented forest mosaic in Andalucía, Southern Spain. Circle size is proportional to patch area in logarithmic scale. Red circles represent patches inhabited simultaneously by the four plant species studied here and constitute the nodes of the networks of spatial genetic variation (Fig. 2). Green nodes indicate habitat patches inhabited by at least one of the four plant species.

Results

The total genetic variation for a species inhabiting a fragmented landscape such as the forest islands in Southern Spain can be partitioned into intra- and interpatch components. The distribution of the intrapatch genetic variation (represented by node size in Fig. 2) shows a gradient of heterogeneity between species. *Q. coccifera* shows the highest intrapatch heterogeneity whereas *C. salviifolius* shows the lowest heterogeneity. The interpatch genetic variation (strength of links in Fig. 2 indicates genetic similarity) ranges from 25% for *Q. coccifera* to 12% for *C. salviifolius*. These two components provide the fundamental elements (nodes and links) of the networks of genetic variation with which we develop our subsequent analysis.

The structure of the networks of genetic variation for the four plant species appears quite similar when considering global pairwise descriptors, such as network connectance (number of established links over all possible links) or number of links per patch. The connectance of the networks of genetic variation is 0.356, 0.352, 0.352, and 0.312 for *C. salviifolius*, *M. communis*, *P. lentiscus*, and *Q. coccifera*, respectively. This reflects that the genetic covariance of each species is sufficiently explained by a similar number of pairs of patches genetically related, slightly lower for *Q. coccifera*. The lower the number of links, the higher the genetic variation between patches.

The cumulative distribution of the number of patches genetically similar to a given patch is best fit to an exponential function in the four cases ($F_{1,6} = 74.272$, $R^2 = 0.925$ for *C. salviifolius*; $F_{1,8} = 35.573$, $R^2 = 0.815$ for *M. communis*; $F_{1,9} = 59.529$, $R^2 = 0.869$ for *P. lentiscus*; and $F_{1,6} = 18.935$, $R^2 = 0.759$ for *Q. coccifera*; $P < 0.05$ in all cases). So, it seems that the four networks are quite homogeneous in terms of this macroscopic variable. There is a well-defined average number of links per patch in the four plant species, similar to what is expected for a randomly assembled network, which means that populations inhabiting each patch tend to have relevant genetic similarity with the same number of other populations.

The above macroscopic view provides a first step in describing network structure based on total number of links and number of links per node. This summary description makes the pattern of genetic variation appear very homogeneous. A further step toward unraveling the structure of these genetic networks is provided by the modularity analysis, which depicts how the above links are organized among groups of patches. That is, we will now look at the identity of the patches to which a given patch is linked.

The modularity analysis depicts a heterogeneous structure of the networks. Specifically, the network of spatial genetic variation for *C. salviifolius*, *M. communis*, and *P. lentiscus* presented a significantly modular structure ($P = 0.003$, $P < 0.001$, and $P < 0.001$, respectively). These species' average modularity level was 0.458 ± 0.002 SD, 0.558 ± 0.001 SD, and 0.498 ± 0.000 SD, respectively ($n = 100$ replicates of the module-finding algorithm in all cases; see *Materials and Methods* for details). This finding implies that the network of genetic variation for these three species is highly structured in modules, where patches within a module are more genetically similar than patches in different modules (modules are color coded in Fig. 2). Therefore, genetic variation is not uniformly distributed, but aggregated in modules. These modules are a bottom-up classification of genetically meaningful units (i.e., a surrogate for real populations). Therefore, our network analysis depicts the relevant scales at which genetic variation is organized.

The classification of forest patches into modules does not reflect a simple geographic distribution (Fig. 2). Specifically, the average distance between two patches within the same module is not statistically shorter than the average distance between any two patches in the network (*C. salviifolius*, Student's $t = -0.065$, $P = 0.474$, $df = 308$; *M. communis*: $t = 0.222$, $p = 0.588$, $df = 321$; *P. lentiscus*: $t = 0.666$, $P = 0.747$, $df = 319$; *Q. coccifera*: $t = 0.599$, $P = 0.725$, $df = 313$).

Q. coccifera, on the other hand, did not present a significant modular structure ($P = 0.061$, 0.343 ± 0.002 SD for the real

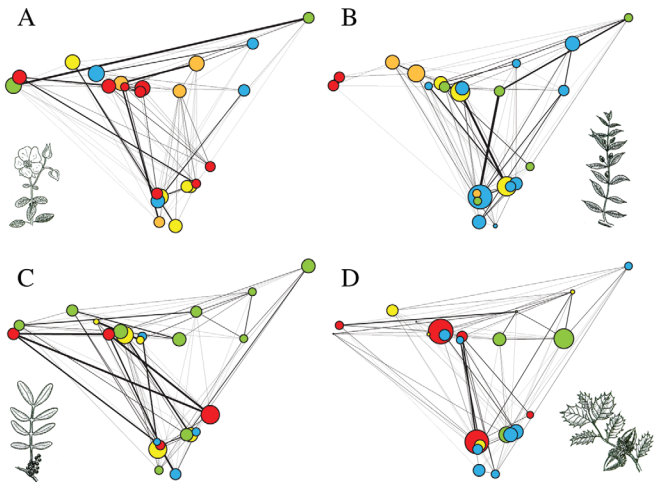

Fig. 2. Networks of spatial genetic variation for the four plant species studied. Here we show one replicate of the module-finding algorithm. (*A*) *Cistus salviifolius*. (*B*) *Myrtus communis*. (*C*) *Pistacia lentiscus*. (*D*) *Quercus coccifera*. Nodes represent habitat patches holding a population of these species. Node position reflects the geographic coordinates of the forest patch whereas node size indicates the intrapopulation genetic variance in relation to the total genetic variance for each species (in linear scale). Links represent significant genetic similarity between pairs of populations once the genetic similarity to other populations has been removed. The pattern of genetic covariance among populations is sufficiently explained for each species by the subset of links here shown. The thickness of the links indicates the level of genetic similarity among populations (same linear scale for all species). Colors represent modules, that is, groups of patches holding genetically similar populations.

ECOLOGY

network, and 0.309 ± 0.022 SD for the population of randomizations), which implies that genetic variation for this species is distributed homogeneously through this fragmented landscape.

Having quantified the overall structure of the networks of spatial genetic variation, we now turn to the role of individual patches within the network. Because, as noted above, a significant modular organization has been found for all species but *Q. coccifera*, we omit the latter in the following analysis, which assumes the existence of a modular organization. Previous analysis of complex networks has identified different roles for nodes in terms of their connectivity both within their module and among modules (11, 15, 16). Specifically, the participation coefficient *PC* indicates how well distributed the links of a node are among different modules *(Materials and Methods)*. Although the bulk of nodes have limited structural importance, a few nodes are extremely important by connecting several such modules (15, 16). The identification of these module connectors will point us toward patches that are disproportionally important for genetic connectivity among modules and thus inform conservation.

If a patch plays a similar role as a connector of modules across all species, we would find a positive and significant correlation between the rank of a patch's participation coefficient through the species-specific networks. This is not the case. Spearman's rank correlation coefficients were not significant ($\rho = 0.234$, $P = 0.283$ for *C. salviifolius-M. communis*; $\rho = -0.133$, $P = 0.546$ for *C. salviifolius-P. lentiscus*; $\rho = 0.029$, $P = 0.897$ for *C. salviifolius-Q. coccifera*; $\rho = -0.063$, $P = 0.776$ for *M. communis-P. lentiscus*; $\rho = 0.528$, $P = 0.010$ for *M. communis-Q. coccifera*; $\rho = -0.059$, $P = 0.789$ for *P. lentiscus-Q. coccifera*). So it seems that the role of each habitat patch in this fragmented landscape is species-specific. Each species here studied will provide a different assessment of the most important patches for the maintenance of genetic connectivity across the network.

Discussion

Potential Processes. A drawback of the results here presented is that they are based on a static description of the spatial pattern of genetic variation. A more challenging task is identifying the processes that generate such patterns. Our study provides a unique scenario to attempt this. Previous across-species comparisons are very constrained by unequal methodology, genetic markers, and study areas. Instead, here we restricted our sampling strategy only to habitat patches inhabited by the four studied species at the time, explicitly assuming that patch histories (e.g., grazing and agriculture) have similarly impacted the four species. Our results are therefore strictly comparable, and differences across species are probably due to their life-history attributes. Furthermore, the species were deliberately selected to represent contrasting life histories (i.e., breeding and seed-dispersal systems). Thus, *C. salviifolius* is hermaphroditic, insect-pollinated, self-incompatible, and barochorous; *M. communis* is hermaphroditic, self-compatible, insect-pollinated, and its berries are actively dispersed by birds and mammals; *Q. coccifera* is monoecious, self-incompatible, wind-pollinated, and its acorns are locally dispersed by small mammals; and *P. lentiscus* is dioecious, wind-pollinated, and its drupes are actively dispersed by birds. These contrasting life histories result in two broad groups of dispersal distances. Thus, whereas *C. salviifolius* and *Q. coccifera* almost certainly have exclusive within-patch dispersal, both *P. lentiscus* and *M. communis* probably experience some between-patch dispersal events. Unfortunately, there is no clear match between these two dispersal groups and the modular versus nonmodular structure of the respective networks of genetic variation. Therefore, we need to turn to other life-history traits.

Differences in Network Structure Between *Q. coccifera* and the Other Three Species. It is difficult to adduce an explanation for the difference in network structure between *Q. coccifera* (nonmodular) and the other three species (modular), but diverse life-history characteristics of this species are potentially at work. *Q. coccifera* has a high capacity of clonal expansion and of formation of large genets. As a consequence, it is quite resistant to being genetically eliminated from a patch. This resistance could explain the high interpatch genetic variation we report here as well as the high allelic and genetic richness at the species level in the study system. This probably means that levels of genetic diversity in *Q. coccifera* are similar to a prefragmented state. However, extensive natural hybridization with the holm oak (17) could be a contributing factor to differences in the network of *Q. coccifera*. Natural hybridization and introgression in plants are indeed sources of evolutionary potential and genetic novelty (18).

Lack of a Spatial Segregation of Modules and Different Roles of Patches. The lack of a geographic concordance of the modules suggests that there is no correlation between geographic and genetic distance, a result congruent with additional analysis showing that there is no regional equilibrium between gene flow and genetic drift in the four species (only marginally for *C. salviifolius*). This result, together with the lack of concordance in the identity of connector patches across species, also suggests that the different species perceive the landscape differently. Thus, life-history traits affect how species perceive their landscape, which is consistent with recent evidence that the mating system of species influences the genetic structure of their populations (19).

Another potential explanation for the lack of geographic concordance of the modules would be that current patterns of genetic variability better reflect past landscape properties than current ones. This hypothesis is supported by two facts. First, this landscape has been greatly transformed in recent times. Specifically, in the last fifty years, a focal patch in this study has lost an average of four neighboring forest fragments in a 500-hectare buffer (the distribution ranging from a loss of 17 patches to a net gain of one patch). Second, the genetic markers used, allozymes, are better indicators of past large events than of small-scale recent and current events. This evidence would reflect a situation in which recent land transformation has not reached a new equilibrium, a likely situation in this type of Mediterranean landscape.

Conservation Implications. Conservation has traditionally been based on single- or multiple-species strategies where species are the explicit targets (20, 21). Our across-species approach, in detecting the existence of genetic modules and species-specific responses to fragmentation, supports the view that not only species but also idiosyncratic processes of capital importance in plants, such as pollen and seed gene flow, deserve detailed attention by researchers and managers (22). We believe that, compared with traditional population summary statistics, our network approach captures the true interpopulation complexity existing in nature and is a starting point for the conservation of biodiversity as a whole. The integrated use of population graphs and modularity analysis shown differently allows a rigorous, bottom-up identification of (*i*) the spatial scale or conservation unit (e.g., a patch, a module with several patches, or the entire network) and (*ii*) the most important habitat patches for the connectivity of the entire landscape.

Regarding point (*i*) above, evolutionarily significant units or management units have been widely discussed in conservation genetics. Although methods dealing with continuous genetic variability within populations are difficult to implement, methods based on discrete genetic units are more easily handled (23). Our modularity approach defines discrete evolutionary units—the modules—that are amenable to incorporation in conservation planning.

Point (*ii*), namely the identification of patches acting as among-module connectors, may be very useful when prioritizing conservation effects. Such connectors do not need to be very well-connected patches but rather patches connected to other

patches from different modules, information that requires a network approach to obtain. These patches play an important role in maintaining the pattern of genetic variability across the entire landscape. Our modularity approach has not been discussed differently as a conservation tool in fragmented habitats (see, however, ref. 12) although it has been discussed in relation to networks of species interactions (24, 25).

Importantly, although the identification of modules as conservation units could be performed in three out of four species—for which the underlying modular structure of genetic variability is significant—the specific ranking of habitat patches is different across the three species. This difference may represent a serious challenge in the conservation of multispecies assemblages. We need additional studies to assess how general this result is and, if so, how we can come up with novel techniques to overcome these difficulties. For example, the methods illustrated here can serve to generate a population of habitat patches, all acting as module connectors for one or a few species. One could then concentrate on this small number of critical connector patches even though they are different for the different species.

Conclusion. To sum up, we have compared, for the first time, the structure of genetic variation across different species inhabiting the same landscape. We have found a common pattern of modular organization in three out of four species but also an independent ranking of patches from the point of view of their role as connectors of different modules. Our paper is a step toward a study of metawebs defined as the collection of networks of genetic variation of all species within a community. Quantifying the variation across such a metaweb will inform us about what properties are general across groups of species and what properties are species-specific. A network approach may contribute to quantifying the consequences of habitat fragmentation for the persistence of genetic variability and to finding critical destruction values beyond which there is a substantial loss of genetic variability and therefore a limit on adaptation to changing conditions.

Materials and Methods

Study Area. The study area is the Guadalquivir River Valley, an area of 21,000 km^2 in Western Andalucía, Southern Spain (see Fig. 1). This area is a fertile countryside with a flat orography ranging in altitude between sea level and 200 m. The climate is Mediterranean, with warm, dry summers and cool, humid winters. Although virtually eliminated from the area, the esclerophylous Mediterranean maquis associated with *Quercus suber L*. and *Quercus ilex*, subsp. *ballota* (Desfontaines; Sampaio) is native to the entire region. However, disclimatic plantations of stone pine (*Pinus pinea L*.) dating back to the eighteenth century are extensive in the area and have become representatives of seminatural vegetation.

Across the Guadalquivir River Valley, 535 forest patches were located and inventoried (26), totalling a surface area of 22,931.5 hectares. The patch area oscillated between 0.19 and 1,737 hectares; but mean (± SD) and median values of the frequency distribution were 42.86 ± 102 and 12.3 hectares, respectively. Mean (± SD) woody plant-species richness at the patch level was 13.4 ± 7.1 (range 1-38). The most frequently recorded species were *Asparagus* spp., *Cistus* spp., *Daphne gnidium*, *Chamaerops humilis, Pistacia lentiscus, Halimium halimifolium, Lavandula stoechas, Olea europaea, Myrtus communis, Quercus coccifera, Phlomis purpurea,* and *Retama sphaerocarpa*.

Molecular Data. To study the spatial variation of the genetic structure in our four plant species, we used isozymes as multivariate codominant markers extracted from young leaves and developed following the standard procedures described in Weeden and Wendel (27) and Soltis et al. (28).

The networks of genetic variation analyzed are based on data from 2,559 individual plants (*Cistus*, 678; *Myrtus*, 662; *Pistacia*, 655; *Quercus*, 564) collected in 23 hard-edges forest patches where the four species coexist. The number of detected loci was 13, 12, 11, and 10 for *Cistus*, *Myrtus*, *Pistacia*, and *Quercus*, respectively. The total number of alleles (and allele range per loci) was 29 (1-5), 22 (1-4), 23 (1-5), and 42 (1-10), for each species, respectively (see *SI Text* for a detailed information about the enzyme systems successfully stained).

Networks of Spatial Genetic Variation. The conditionally independent network of genetic variation can be represented algebraically by an incidence matrix *A*, in which each element a_{ij} denotes the presence (nonzero value) or absence (zero value) of genetic similarity connecting populations *i* and *j*. The higher the value of the link, the higher the conditional dependence of the genetic covariance between the pair of linked populations.

The main steps for calculating the network of spatial genetic variation of a species are (*i*) calculating the genetic distance between populations by translating multilocus genotypes of individuals to multivariate codification vectors and (*ii*) estimating the conditional independence structure of the genetic covariance.

Calculation of the Genetic Distance Between Populations by Translating Multilocus Genotypes to Multivariate Codification Vectors. We begin by defining the genetic distance between a pair of individuals of the same diploid species for a multiallelic codominant locus, which would be the case with either allozymes or microsatellite (SRR) markers. Following Smouse and Peakall (29), we use an additive scoring system to translate the genotype of an individual into a codification vector *Y* of length *K*, where *K* is the number of *k* alleles in the population. The *y* values of the codification vector for each individual (from $k = 1$ to $k = K$) can be 0, 1, and 2, depending on whether the individual has zero, one, or two copies of the *k* allele (see *Materials and Methods* in the *SI Appendix* for an example of the *Y* vectors corresponding to the three possible genotypes for a locus *A* with two alleles *A*1 and *A*2).

The squared distance between any two individuals with genotypes *i* and *j* is one-half the Euclidean distance between their respective vectors y_i and y_j:

$$d_{ij}^2 = \frac{1}{2}\sum_{k=1}^{K}(y_{ik} - y_{jk})^2. \quad [1]$$

The distance values between inviduals of a diploid species range from zero to two. In the case of two individuals with genotypes *A*1*A*1 and *A*1*A*2, the squared genetic distance between them is

$$d^2 = \frac{1}{2}\left[(2-1)^2 + (0-1)^2\right] = 1. \quad [2]$$

We can extend the codification vector *Y* and the calculation of the genetic distance to *L* loci. Multilocus genotypes are now translated into multivariate coding vectors of a length equal to the number of independently assorting *k* alleles across all *L* loci. See *Material and Methods* in the *SI Appendix* for an example of the *Y* vectors of length equal to 5 corresponding to the 18 possible genotypes for two locus, *A* and *B*, with two (*A*1,*A*2) and three (*B*1,*B*2,*B*3) alleles, respectively.

Therefore, the squared genetic distance between, for example, two individuals with genotypes *A*1*A*1, *B*1*B*2 and *A*1*A*2, *B*3*B*3 is

$$d^2 = \frac{1}{2}\left[(2-1)^2 + (0-1)^2 + (1-0)^2 + (1-0)^2 + (0-2)^2\right] = 4. \quad [3]$$

Let us now move from individuals to *N* populations. We calculate the average genetic individual (centroid) for each *n* population by averaging the multivariate coding vectors of all individuals belonging to the *n* population. The resulting vector for each *n* population is then used to estimate the genetic distance between all pairs of populations following Eq. **1**.

Note that rare alleles are more important for differentiating individuals than are common alleles (30) and should thus be weighted differentially. We can take this fact into account by incorporating the allelic frequency in the calculation of the genetic distances. Following Smouse and Peakal (29), Eq. **1** can be extended to

$$d_{ij}^2 = \frac{1}{2}\sum_{k=1}^{K}\left[\frac{1}{Kp_k}(y_{ik} - y_{jk})^2\right], \quad [4]$$

where the allele-specific weights are inversely proportional to the allelic frequencies p_k and the total number of alleles *K*. Note that for equiprobable alleles we obtain Eq. **1**.

The contribution to the overall genetic variation due to differences among all pairs of populations defines a distance matrix, *D*, whose off-diagonal elements, d_{ij}, represent the statistical distance between the average genetic individual of each pair of populations.

Estimation of the Conditional Independence Structure of the Genetic Covariance. The resulting distance matrix *D* is a fully connected matrix in which all populations are connected to all others by links of weight d_{ij}. The topology of this matrix does not give us information about the interpopulation relationships. The translation of the population distance matrix *D* to a minimal incidence matrix containing the smallest link set that sufficiently describes the genetic covariance structure among populations relies upon the techniques of conditional independence (10).

The next task is, therefore, to identify links that are redundant in describing the simplest network encapsulating the total genetic covariance structure among populations. These genetic relationships can be removed from the network without significantly decreasing the fit of the network of spatial genetic variation to the population genetic data. We used the method of edge deviance to calculate conditional independence, as has recently been described in an evolutionary context by Magwene (31) and followed by Dyer and Nason (10) in an ecological context. The first step is translating the

distance matrix D to a covariance matrix C. Following Gower's (32) transformation and Dyer and Nason's (10) notation, the covariance between populations i and j is

$$c_{ij} = \frac{1}{2}(d_{ij} - d_{i.} - d_{.j} + d_{..}), \quad [5]$$

where the subscripts i and j index the elements of D, and the period subscript "." indexes the mean of the row(s) and/or column(s) in D. Next, we invert the covariance matrix producing a generalized inverse matrix called a precision matrix P. (33). If an element of the precision matrix is zero, the corresponding populations are conditionally independent given the remaining populations. Each diagonal element is related to the multiple correlation coefficient R_i^2 between population i and the remaining populations: $p_{ii} = 1/(1 - R_i^2)$, which is a measure of the proportion of the genetic variation in the ith population jointly accounted by the remaining populations. After that, we scale the precision matrix so that the main diagonal is composed of ones, and the off-diagonal partial correlation coefficients between i and j are given by

$$r_{ij} = \frac{-p_{ij}}{\sqrt{(p_{ii}\ p_{jj})}}. \quad [6]$$

By changing the sign of the off-diagonal elements, we obtain the correlation matrix R.

As in the precision matrix, absolute values of r_{ij} which are zero denote pairs of populations whose covariance structure is conditionally independent given all the other populations. Finally, the estimation of how small an element r_{ij} must be to be considered zero is based on the statistic called edge exclusion deviance (EED) (τ) described by Whittaker (34):

$$\tau = -I\ Ln[1 - (r_{ij})^2], \quad [7]$$

where I is the number of individuals in the entire dataset. The EED is an information theoretic measure, with an asymptotic χ^2 distribution, of whether a particular link can be eliminated from the fully connected correlation matrix R. Each EED value tests a single link. The value of each r_{ij} element is tested against the χ^2 distribution with one degree of freedom. All r_{ij} values with deviances less than 3.84 (the 5% threshold of the χ^2 distribution with $df=1$) are rejected (31, 34). This means that the r_{ij} values of those links are not significantly higher than zero, and thus those pairs of populations are conditionally independent. This provides the minimum number of links that explain the overall pattern of population genetic covariation.

The EED is based on the concept of information divergence (34). This concept can also provide the strength of the links, that is, how strong is the genetic correlation between any pair of connected populations. This strength is measured by the information of population i about population j and vice versa, conditional on all the remaining populations. For any pair of connected populations, the strength of the genetic dependence is calculated as

$$s_{ij} = -\frac{1}{2}\ Ln[1 - (r_{ij})^2]. \quad [8]$$

Note that the strength of the link is zero when the partial correlation r_{ij} is zero.

In summary, the network of spatial genetic variation of a species is created by: (*i*) translating multilocus genotypes of individuals to multivariate codification vectors; (*ii*) estimating genetic distances between populations from these codification vectors taking into account the allelic frequencies; (*iii*) translating the genetic distance matrix to a covariance matrix; (*iv*) inverting the covariance matrix to obtain a precision matrix; (*v*) standardizing the precision matrix to a correlation matrix; (*vi*) estimating the conditional independence structure of the genetic covariance using the edge exclusion deviance; and (*vii*) calculating the strength of the genetic dependence between populations. A working example of these steps is illustrated in *Material and Methods* in the *SI Appendix*.

The goodness of fit for the resulting topology of the network of spatial genetic variation can be evaluated analytically by estimating the model deviance (10, 31, 34). EEDs alone cannot be sufficient to specify a final network with adequate fit (34). The addition of links in a new network may improve the fit of the model, that is, a smaller deviance of a new network can fit sufficiently well relative to the fully connected network. The deviance difference between the new network and the previous one can be significant. Even if more connected networks fit slightly better, the resulting topological patterns will remain likely unaltered.

Modularity Analysis. We used a module-detection algorithm (35) combined with a simulated annealing optimization approach (15,16) to detect high-level population modules. Specifically, we have used the simplest generalization to weighted networks of the modularity implemented in Guimerà and Amaral's algorithm (36) The algorithm follows a heuristic procedure to find an optimal solution for the maximization of a function called modularity (35). For weighted networks the modularity is given by (36)

$$M_W(P) = \sum_{s=1}^{N_M}\left[\left(\frac{w_s^{in}}{W}\right) - \left(\frac{w_s^{all}}{2W}\right)^2\right], \quad [9]$$

where, $W = \sum_{i \geq j} w_{ij}$, w_s^{in} is the sum of the weights of the links w_{ij} within module s, and $w_s^{all} = \sum_{i \in s}\sum_j w_{ij}$.

Optimization of this function maximizes the weights of genetic dependences between populations belonging to the same module and minimizes the weight of genetic dependences between populations belonging to different modules. In a network with high modularity, the density of links (and their weights) inside modules is significantly higher than the random expectation. Because the detection of the modularity is a heuristic process, we run 100 replicates of the simulated annealing algorithm for each plant species. From these analyses we obtained the average value of modularity and the average number and identity of modules detected by the algorithm. We also estimated how well distributed the genetic dependences of a patch are among different modules (participation coefficient, varying between 0 and 1). This allows us to estimate the role of each population as connectors of genetic variation between modules across the landscape (see details in refs. 15, 16).

To assess the significance of this modular structure, we compared the modularity level with that corresponding to 1, 000 randomizations of the network for each species, preserving the number of links per patch.

The number of genetic modules is 5.100 ± 0.345 SD, 5 ± 0.000 SD, and 4 ± 0.000 SD, respectively. So there was almost no variation across the 100 replicates (the module-finding algorithm always ended up detecting the same number of modules for Myrtus and Pistacia), whereas for Cistus this fraction is 0.9.

To assess the consistence of the results given by the modularity algorithm, we quantified how conserved the distribution of patches within modules was across replicates. We calculated, for a given pair of patches observed in the same module in one replicate, how often that particular pair of patches was also classified within the same module in the remaining 99 replicates. Coincident results represent 0.93%, 0.92%, and 0.99% of the cases, respectively). We chose one replicate with a consistence across replicates equal to the average for the representation of the network of spatial genetic variation for each species (modules are color coded in Fig. 2).

ACKNOWLEDGMENTS. We thank Rodney Dyer for his help on methodological questions and Carlos J. Melián, Daniel B. Stouffer, and Jason Tylianakis for useful comments on a previous version of this paper. This work was funded by the Junta de Andalucía through the Excellence Grant P06-RNM-01499 (to A.A.), a PhD Fellowship from the Spanish Ministry of Education and Science (to M.A.F.), and the European Heads of Research Councils, the European Science Foundation, and the EC Sixth Framework Program through a European Young Investigator Award (to J.B.).

1. Hanski I (1999) *Metapopulation Ecology* (Oxford Univ Press, New York).
2. Hanski I, Gaggiotti OE (2004) *Ecology, Genetics, and Evolution of Metapopulations* (Elsevier-Academic, London).
3. Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford).
4. Haldane J (1932) *The Causes of Evolution* (Longmans Green, London).
5. Malécot G (1969) *The Mathematics of Heredity* (WH Freeman, San Francisco).
6. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97-159.
7. Wright S (1943) Isolation by distance. *Genetics* 28:114-138.
8. Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends Ecol Evol* 18:189-197.
9. Storfer, et al. (2007) Putting the landscape in landscape genetics. *Heredity* 98:128-142.
10. Dyer RJ, Nason JD (2004) Population graphs: The graph theoretic shape of genetic structure. *Mol Ecol* 13:1713-1727.
11. Fortuna MA, García C, Guimarães PR, Bascompte J (2008) Spatial mating networks in insect-pollinated plants. *Ecol Lett* 11:490-498.
12. Garroway CJ, Bowman J, Carr D, Wilson PJ (2008) Applications of graph theory to landscape genetics. *Evol Appl* 1:620-630.
13. Rozenfeld, et al. (2007) Spectrum of genetic diversity and networks of clonal organisms. *J R Soc Interface* 4:1093-1102.
14. Rozenfeld, et al. (2008) Network analysis identifies weak and strong links in a metapopulation system. *Proc Natl Acad Sci USA* 105:18824-18829.
15. Guimerà R, Amaral LAN (2005) Cartography of complex networks: Modules and universal roles. *J Stat Mech Theory Exp* 1:P02001.
16. Guimerà R, Amaral LAN (2005) Functional cartography of complex metabolic networks. *Nature* 433:895-900.
17. Rubio de Casas, et al. (2007) Taxonomic identity of *Quercus coccifera* L. in the Iberian Peninsula is maintained in spite of widespread hybridization, as revealed by morphological, ISSR, and ITS sequence data. *Flora* 202:488-499.
18. Rieseberg LH (1997) Hybrid origins of plant species. *Ann Rev Ecol Syst* 28:359-389.
19. Duminil J, et al. (2007) Can population structure be predicted from life-history traits? *Am Nat* 169:662-672.
20. Lambeck RJ (1997) Focal species: A multispecies umbrella for nature conservation. *Cons Biol* 11:849-856.
21. McCarthy MA, Thompson CJ, Williams NSG (2006) Logic for designing reserves for multiple species. *Am Nat* 167:717-727.
22. Thrall PH, Burdon JJ, Murray BR (2000) The metapopulation paradigm: A fragmented view of conservation biology. In *Genetics, Demography and Viability of Fragmented Populations*, eds Young AG, Clarke GM (Cambridge Univ Press, Cambridge, UK).
23. Diniz–Filho JAF, Telles MPC (2006) Optimization procedures for establishing reserve networks for biodiversity conservation taking into account population genetic structure. *Gen Mol Biol* 29:207-216.
24. Olesen JM, Bascompte J, Dupont YL, Jordano P (2007) The modularity of pollination networks. *Proc Natl Acad Sci USA* 104:19891-19896.
25. Rezende E, Albert EM, Fortuna MA, Bascompte J (2009) Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecol Lett* 12:779-788.

26. Aparicio A (2008) Descriptive analysis of the 'relictual' Mediterranean landscape in the Guadalquivir River valley (southern Spain): A baseline for scientific research and the development of conservation action plans. *Biodivers Conserv* 17: 2219–2232.
27. Weeden NF, Wendel JF (1989) Visualization and interpretation of plant isozymes. In *Isozymes in Plant Biology*, eds Soltis DE, Soltis PE (Chapman & Hall, London).
28. Soltis DE, Haufler CH, Darrow DC, Gastony GE (1983) Starch gel electrophoresis or ferns: A compilation of grinding buffers, gel and electrode buffers, and staining schedules. *Am Fern J* 73:9-27.
29. Smouse PE, Peakal R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82:561-573.
30. Epperson BK (1995) Fine-scale spatial structure: Correlations for individual genotypes differ from those for local gene frequencies. *Evolution* 49:1022-1026.
31. Magwene PM (2001) New tools for studying integration and modularity. *Evolution* 55:1734-1745.
32. Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325-338.
33. Cox JM, Wermuth N (1996) *Multivariate Dependencies: Models, Analysis, and Interpretations* (Chapman & Hall, New York).
34. Whittaker J (1990) *Graphical Methods in Applied Multivariate Statistics* (Wiley, New York).
35. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113.
36. Guimerà R, Sales-Pardo M, Amaral LAN (2007) Module identification in bipartite and directed networks. *Phys Rev E* 76:036102.