Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes

Ilkka Hanski\(^1\), Torsti Schulz\(^1\), Swee Chong Wong\(^1\), Virpi Ahola\(^1\), Annukka Ruokolainen\(^1\) & Sami P. Ojanen\(^1\)

Ecologists are challenged to construct models of the biological consequences of habitat loss and fragmentation. Here, we use a metapopulation model to predict the distribution of the Glanville fritillary butterfly during 22 years across a large heterogeneous landscape with 4,415 small dry meadows. The majority (74%) of the 125 networks into which the meadows were clustered are below the extinction threshold for long-term persistence. Among the 33 networks above the threshold, spatial configuration and habitat quality rather than the pooled habitat area predict metapopulation size and persistence, but additionally allelic variation in a SNP in the gene Phosphoglucose isomerase (Pgi) explains 30% of variation in metapopulation size. The Pgi genotypes are associated with dispersal rate and hence with colonizations and extinctions. Associations between Pgi genotypes, population turnover and metapopulation size reflect eco-evolutionary dynamics, which may be a common feature in species inhabiting patch networks with unstable local dynamics.
Habitat loss and fragmentation are the main drivers of ongoing loss of biodiversity\(^1\)–\(^3\), but ecologists have made only limited progress in predicting the consequences of habitat loss on population viability and extinction\(^4\)–\(^6\) in quantitative terms, and in clarifying the additional effects of fragmentation\(^6\)–\(^7\). It has been suggested that framing the question as a dichotomy between the effects of loss of pooled habitat area versus fragmentation per se may be misleading, and that we should instead start by considering the causal mechanisms that underpin species’ responses to altered spatial structure of habitat\(^8\)–\(^9\). One approach that does this but is applicable only to highly fragmented landscapes (defined below) is based on the measure of metapopulation capacity, which integrates the effects of habitat amount and spatial configuration (fragmentation) into a single number.

From the viewpoint of practical conservation, it would be critically important to be able to answer questions such as how much habitat is sufficient for long-term persistence of populations and species. The species-area relationship (SAR) (ref. 10) has been used to predict extinction due to habitat loss\(^11\)–\(^14\), but these predictions continue to be debated\(^1\)–\(^3\). They only provide a rough baseline prediction at best, because SAR does not take into account the temporal population dynamics following habitat loss, and SARs completely ignore any fragmentation effects (but see refs 16–18). In brief, ecologists are challenged to produce more predictive models of the consequences of habitat loss and fragmentation. Given the great variation in the ecological circumstances in which species and communities occur, and the range of spatial scales from small experimental study plots\(^19\) to the continental\(^20\) and even global scale\(^21\), it might be advisable to develop analyses and models for more circumscribed situations rather than aiming for universal approaches.

One common situation in nature is represented by highly fragmented landscapes, in which the focal habitat accounts for only a small percentage of the total landscape area, of the order of 1%. Very large numbers of invertebrates, plants and fungi inhabit highly fragmented landscapes\(^22\), and human land use makes the environment of large numbers of yet other species increasingly fragmented\(^23\)–\(^24\). In highly fragmented landscapes, the focal habitat occurs in small patches, or fragments, which are distributed in a more or less aggregated manner across the landscape. The classical metapopulation approach\(^25\), assuming frequent extinctions of small local populations and frequent establishment of new populations in the currently unoccupied patches, has been developed for species inhabiting highly fragmented landscapes.

The large metapopulation of the Glanville fritillary butterfly (Melitaea cinxia) in the Åland Islands in Finland is a model system for the study of the ecological, genetic and evolutionary consequences of habitat fragmentation\(^25\)–\(^26\). The landscape consists of a very large network of dry meadows with the pooled habitat area covering about 1% of the landscape\(^27\). Here, we analyse data for 22 years, comprising 66,527 records of the presence or absence of the butterfly in 4,415 habitat patches distributed among 125 semi-independent networks. Our aim is to test the predictive power of a spatially realistic metapopulation model\(^28\) and to examine other factors apart from habitat area and fragmentation that may influence metapopulation size and persistence. In particular, we analyse the effects of habitat quality, and the effect of a well-studied candidate gene, Phosphoglucose isomerase (Pgi), on the butterfly metapopulation dynamics and metapopulation size.

We show that the majority of the habitat networks of the butterfly are below the extinction threshold. Metapopulation persistence and sizes of the networks above the extinction threshold can be predicted by spatial configuration, habitat quality and Pgi genotypes. In accordance with earlier work in other systems\(^29\)–\(^30\), associations between Pgi genotypes, population turnover and metapopulation size show compelling evidence that demographic and genetic dynamics are closely coupled, leading to observable eco-evolutionary dynamics in real systems at landscape scales.

**Results**

**Extinction threshold in fragmented landscapes.** The study system covers an area of 50 × 70 km (Fig. 1a), and consists of a large network of 4,415 dry meadows with one or both of the two host plant species of the butterfly\(^27\). The number of meadows that have been known to us and surveyed since the beginning of the study in 1993 has varied. Most importantly, the entire study area was re-mapped for the habitat in 1998–1999, which greatly increased the number of known meadows. For the analyses in which missing data would greatly affect the result, we have used the data set for the years 1999–2014.

Control surveys from the years 2009, 2011 and 2015 show that the presence of the butterfly is not detected in up to 15% of occupied meadows, but the non-detection only concerns meadows where the population is very small (Supplementary Tables 1–3). Almost half of the 29 non-detected populations consist of a single larval group while the mean size of the detected populations is eight. Thus the populations that are missed have a very small influence on the dynamics of the metapopulation as a whole.

There is much spatial variation in the density of meadows across the study landscape. The life-time of the butterfly are mostly limited to 2–3 km\(^31\). This means that, in any one generation butterflies mostly move mate and reproduce within areas that are less than 1% of the total landscape area. We have clustered the habitat patches into 125 semi-independent networks (Fig. 1a), which differ in terms of the number and spatial configuration of the patches and thus offer an opportunity to analyse the effects of landscape structure on the dynamics and distribution of species. Metapopulations inhabiting these networks are dynamically relatively independent from each other, though dispersal does occur between the networks and affects the dynamics of some of them, as will be demonstrated below. In the analyses, we use networks and the respective metapopulations as independent data points.

The variation in the number of patches in a network, and variation in the total amount of habitat (pooled area of patches) in a network explain only 9 and 11% of variation in the fraction of patches occupied, averaged across all years and denoted by \(p\) (Supplementary Figs 1 and 2). However, the fraction of patches occupied (\(p\)) is not a powerful measure of regional abundance, because it gives equal weight to small and large, and to well-connected and isolated habitat patches, which play very different roles in the dynamics of the metapopulation and hence have dissimilar consequences for long-term abundance\(^32\). Metapopulation theory\(^28\) suggests an alternative measure, \(p_\delta\), which is a weighted average of patch occupancies, where the weights describe the role of individual patches in the dynamics of the metapopulation. In the deterministic model of patch occupancy dynamics\(^33\), the equilibrium value of \(p_\delta\) is given by

\[
\hat{p}_\delta = 1 - \frac{\delta}{\lambda_M},
\]  

where \(\delta = e/c\) is the ratio of the extinction and colonization rate parameters and is called the extinction threshold. We emphasize that the extinction threshold is a characteristic of the species, and hence the metapopulations in different networks are assumed to
have the same value. $\lambda_M$ is called the metapopulation capacity. This measure describes the features of the networks, and hence different networks in this study have different values. Mathematically, $\lambda_M$ is the leading eigenvalue of matrix $M$, the elements of which describe patch-specific extinction and colonization rates. Metapopulation capacity integrates the effects of patch areas and their spatial locations on the capacity of the network to support a viable metapopulation.  

To apply equation (1) to data, we need to calculate the value of the metapopulation capacity ($\lambda_M$) for each network, and to estimate the extinction threshold ($\delta$) for the species. For the latter, there are two approaches, we may use an expression that gives the incidence of occupancy for each habitat patch at equilibrium, or we may use empirical data on observed extinctions and colonizations. The first approach is based on the equation giving the occupancy of patch $i$ at equilibrium as $\hat{p}_i = C_i / (C_i + E_i)$, where $C_i$ and $E_i$ are the colonization and extinction rates. We make standard assumptions about how landscape structure affects $C_i$ and $E_i$, and add habitat quality in the model as explained in the Methods. We thereby derive the following equation for $\hat{p}_i$

$$\hat{p}_i = \frac{1}{1 + \frac{1}{M^2}}$$

where $\hat{p}_i$ is the fraction of the 22 years that patch $i$ has been occupied, the variables $A_n$, $Q$, and $S_i$ are the area, quality and connectivity of patch $i$, respectively, and $x$ and $y$ are two

---

**Figure 1 | The study system and the metapopulation capacity of the habitat patch networks.** (a) The red polygons demarcate habitat patches in the 33 networks above the extinction threshold ($\delta = 5.47$), while the 92 light-blue polygons are networks below the threshold (the smallest ones are not visible). The map contains shoreline data from the National Land Survey of Finland Topographic map 1:100,000 02/2015. Scale bar is 10 km. (b) Metapopulation size ($\hat{p}_i$) in each network as a function of metapopulation capacity ($\lambda_M$). The continuous line gives metapopulation size as predicted by equation (1) parameterized with patch-level data using equation (2). Networks above and below the extinction threshold are shown by red and light-blue dots, respectively ($n = 125$). (c-e) The three panels give time series of metapopulation size for three networks, one of which (c) is above the threshold, one is the island Sottunga (d) and the third one is a network below the threshold (e). The value of the metapopulation capacity is given for each network. The two lines are the fraction of occupied patches (thin line) and $\hat{p}_i$ (thick line), which gives more weight to the dynamically more important patches. Sottunga was unoccupied in 1991, in which year the butterfly was translocated there. (f) A ‘winter nest’ span by fifth instar larvae at the base of the host plant (P. lanceolata). The winter nests, inside which the larvae diapause, are sufficiently conspicuous to make the large-scale census of populations feasible. Photograph by Sami Ojanen.
Viable versus non-viable patch networks. We next use equation (1) to calculate $\hat{p}_1$ for the 125 networks into which the total of 4,415 habitat patches were clustered. In a linear regression, metapopulation capacity explained 40% of variation in the average of the annual $\hat{p}_1$ values ($F$-test: $F_{1,123} = 84.16$, $P < 10^{-14}$), which is substantially more than what the simple measures of network structure (patch number and pooled habitat area) explained of variation in the fraction of occupied patches (ca. 10%, above). More importantly, equation (1) predicts $\hat{p}_1$ as a non-linear function of $\lambda_M$ (Fig. 1b). The model fits the 33 networks above the extinction threshold (for an example see Fig. 1c) reasonably well ($r = 0.42$, $t$-test: $t_{31} = 2.68$, $P = 0.0115$), though there is variation in $\hat{p}_1$, which to a large extent is due to a strong genetic effect (below). For the 92 networks with $\lambda_M < \delta$, equation (1) predicts that the respective metapopulations are not viable, the predicted $\hat{p}_1$ values are negative. Contrary to this prediction, however, many of these metapopulations are fairly large (Fig. 1b), though only temporarily so: 79 of the 92 networks (86%) were extinct (not a single occupied patch) for at least 5 years out of the 22 years (for an example see Fig. 1e). Presence of the butterfly in the ‘non-viable’ networks may be due to dispersal from outside, which may rescue the metapopulation temporarily from permanent extinction. This hypothesis is supported by the present data: connectivity of the focal network to patches in the surrounding networks has a significant effect on metapopulation size ($\hat{p}_1$) in the 92 non-viable networks (Supplementary Table 4).

Among the 33 networks above the extinction threshold, in a linear regression connectivity to the surrounding networks does not explain variation in $\hat{p}_1$ ($R^2 = 0.00$, $F$-test: $F_{1,31} = 0.07$, $P = 0.80$), and hence we conclude that the dynamics of these metapopulations are relatively independent, not much influenced by dispersal from outside. Only 5 of the 33 networks were extinct, or apparently extinct (due to non-detection), for at least 5 years out of the 22 years, which is a much smaller percentage (15%) than in the networks below the threshold (86%) (Pearson’s $r^2$-test: $r^2 = 55.1$, $P < 0.0001$). The island of Sottunga outside the main Åland (Fig. 1a) is an informative example. This island with an area of 9.2 km$^2$ and 49 small patches with the pooled area of 9.1 ha was unoccupied in 1991, when 62 larval groups were translocated to the island from Sottunga outside the main Åland (Fig. 1a) is an informative example. This island with an area of 9.2 km$^2$ and 49 small patches with the pooled area of 9.1 ha was unoccupied in 1991, when 62 larval groups were translocated to the island from
Genetic effects on metapopulation size and persistence. The model above accounts for the effects of the spatial configuration and habitat quality on $p_2$. Previous work on the Glanville fritillary has shown that the single nucleotide polymorphism (SNP) pgic:c.331A>C in the glycolytic gene $P_g$ is significantly associated with flight metabolic rate (FMR) and dispersal rate in the field$^{31,34}$. This is relevant here because dispersal and colonizations necessarily influence $p_2$. We characterize metapopulations in the 33 networks above the extinction threshold by the pooled frequency of the AC and CC individuals in the SNP pgic:c.331A>C, denoted by $f_{disp}$. These butterflies have higher FMR and dispersal rate than the AA homozygotes. We use a large data set sampled across the entire study area in 2007–2012 and genotyped for pgic:c.331A>C as well as for 18 other SNPs (Supplementary Table 5). In the case of pgic:c.331A>C, this material includes a sample from >10 larval family groups per network for 26 of the 33 networks, with a median sample size of 68 individuals per network (Supplementary Data 1). Sample sizes are comparable for the other SNPs (Supplementary Table 5). In these 26 networks, $f_{disp}$ increases with decreasing pooled area of habitat and with increasing rate of population turnover (Fig. 2a). The former is the product of the number of patches and the average patch area, and hence $f_{disp}$ is high in networks with a small number of small patches (Supplementary Fig. 4). The effect of population turnover, the rate of extinctions and colonizations, is consistent with the idea that colonizations select for individuals with high dispersal capacity$^{35}$. The results for the 18 other SNPs show that the significant associations involving pgic:c.331A>C are not due to population structure (Fig. 3a). The other significant SNP in Fig. 3a is from the gene Glucose-6-phosphate 1-dehydrogenase ($G6pd$), located next to $P_g$ in the glycolytic pathway, and which is significantly associated with the pooled area of habitat in the network (Supplementary Fig. 5), though not with population turnover. The major alleles in the two SNPs are negatively correlated at the network level ($r = -0.48$, t-test: $t_{23} = -2.26$, $P = 0.037$).

In the networks above the extinction threshold, metapopulation size $p_2$ is significantly correlated with $f_{disp}$ as well as with metapopulation capacity (Fig. 2b). In quantitative terms, 25% increase in $f_{disp}$ from its median value, which corresponds to one standard deviation of the distribution of $f_{disp}$ among the networks, increases metapopulation size by 17% (assuming the median value of $\lambda_M$). This is a surprisingly large effect and not a result of inflation due to (cryptic) population stratification because none of the other SNPs show inflation (Fig. 3b).

Moreover, $f_{disp}$ is associated with metapopulation persistence: three viable networks (#4, #3 and #107) that were extinct or apparently extinct for at least 5 years had lower $f_{disp}$ than the 27 networks that endured during the full 22 years (t-test: $t_{23} = 2.33$, $P = 0.028$; includes those with a small sample size; genetic data were not available for 3 out of the 33 viable networks; Supplementary Data 1).

The results in Fig. 2a suggest that high population turnover in networks made up of a small number of small patches selects for highly dispersive butterflies (large $f_{disp}$), which in turn increases metapopulation size (Fig. 2b). The latter effect follows from first principles, recalling that long-term metapopulation size is determined by a balance between extinctions and colonizations (equation (1), recall that $\delta = c/c$). The probability of colonization of the currently unoccupied habitat patches increases, and the probability of extinction of the existing populations decreases, with $f_{disp}$ in the surrounding populations from which immigrants arrive. We denote the latter variable by $f_{disp}^{imm}$ (Table 3). High $f_{disp}^{imm}$ increases dispersal rate and thereby the colonization rate, and it increases the rescue effect and hence decreases the extinction rate in existing populations. Note that genetic change and demographic change are closely coupled: high $f_{disp}$ in the network increases colonizations and decreases extinctions (Table 3), while extinctions and colonizations (population turnover) increase network-level $f_{disp}$ (Fig. 2a).

Discussion

Our results demonstrate how a large heterogeneous landscape is a mosaic of 'hot' and 'cold' sections in the sense that parts of
the landscape harbour networks with viable metapopulations (viable networks for short), while the rest consists of non-viable networks in which the species is not expected to persist on the long term. In our case, 39% of the 4,415 habitat patches, and 29% of the pooled habitat area, are located in the cold sections, in networks below the extinction threshold. These networks may be temporarily occupied due to dispersal from the more favourable parts of the landscape, and these networks may thereby function as temporary stepping stones and facilitate the spread of a species across large areas. Classifying heterogeneous landscapes into hot and cold sections is helpful for practical conservation, as knowing the structure of the landscape helps direct conservation measures in a meaningful manner.

Our work shows how the landscape classification can be done in practice. To calculate metapopulation capacity, one needs to know habitat patch areas, connectivities and the average dispersal distance of the species. Additionally, further information is needed to estimate the scaling parameters influencing emigration, immigration and extinction rates. If such information is lacking, one may use values for other comparable species and knowledge about the biology of the focal species. One should note that while the value of the metapopulation capacity depends on these parameters, the ranking order of different networks is much less sensitive; and often it is useful to be able to rank networks from the most to the least viable. To give an idea of the empirical features of the 33 networks above the extinction threshold in the present study, they have minimally 32 habitat patches, minimally around 10 ha of habitat, and they minimally cover an area of 5 km² (Supplementary Data 1). Though these figures do not apply to all insect species, they may guide to the habitat requirements of other insect species of intermediate mobility that live in comparable habitat.

At the level of individual habitat patches, the probability of patch occupancy was best explained by patch areas and spatial locations, which are the first order effects in stochastic patch occupancy models. However, several features of habitat quality also made a substantial effect. This is to be expected, as there is always variation of quality in natural habitats across large areas. The result also depends on how the patches have been delineated in the first place. For instance, some potential patches may not be considered as patches at all because they are deemed to have such low quality. If they were nonetheless included among the habitat patches, the effect of habitat quality among the set of patches would be greater.

The numbers of extinction and colonization events out of possible events for the years 1999–2014 are 3,641/9,096 and 3,581/40,438, respectively. The explanatory variables are scaled patch area \( f_{\text{immig}} \) and \( f_{\text{disp}} \). The intercept \( n_0 \) is the estimated scale of annual network effects and \( n_0 \) is the estimated scale of network-specific patch effects. Thus \( r \) and \( \rho \) determine the magnitude and variance of the patch and network level random effects, respectively.
The amount of host plants and the percentages of dry and low vegetation had a positive effect on patch occupancy, while the percentage of grazed patch area had a negative effect. Desiccation of host plants, especially during the early larval instars in July, increases larval mortality and may even lead to local extinction. On the other hand, dry years are beneficial in the long term because they reduce the growth of grasses, which compete strongly with the host plants. Grazing has similar conflicting short-term and long-term effects: grazing (trampling) has a direct negative effect on larval survival, but grazing also maintains habitat quality by preventing plant secondary succession. There have been changes in summer temperature, precipitation and numbers of cattle, horses and sheep in the Åland Islands during the study period, but these changes have not led to an increasing, nor decreasing, trend in the total metapopulation size of the Glanville fritillary. On the other hand, the amplitude of fluctuations in the metapopulation as a whole has increased, reflecting increasing strength of regionally correlated environmental stochasticity in the dynamics. These long-term changes have been attributed to increasing frequency of dry summers, a consequence of climate warming.

The best predictive model at the network level has the metapopulation capacity as the explanatory variable describing the structure of the environment. This measure integrates the effects of the amount and spatial configuration of habitat, and hence there is no need (nor opportunity) to isolate the effects of the two. However, given that the total amount of habitat explains only 10% of variation in metapopulation size, while the metapopulation capacity explains 40%, it is clear that within the range considered the amount of habitat alone has little predictive power and that the spatial configuration (fragmentation) has an important effect. This is consistent with the general notion that fragmentation effects are increasingly important when the total amount of habitat in the landscape is small (here only about 1%)\(^7,38\).

An SNP in the gene \(Pgi\) has a surprisingly large effect in metapopulations inhabiting the viable networks, explaining 30% of variation in metapopulation size. This is a very large single-gene effect, but in qualitative terms the effect is expected based on previous results on \(Pgi\) in this species. The AC heterozygotes have higher FMR than the AA homozygotes by up to 50%, though difference between the genotypes is affected by interaction with body size and ambient temperature.\(^{31,34,39}\) Butterflies with higher FMR have higher dispersal rate in the field.\(^{34}\) More dispersive butterflies can be expected to be better colonizers, for which there are two types of evidence. First, results in Table 3 show directly that the colonization rate is increased by high frequency of \(f_{\text{disp}}\) among the butterflies in the surrounding source populations. Second, the dispersive butterflies are significantly more frequent in newly established than old local populations.\(^{35}\) This comparison involves the F1 offspring of the actual colonizers, but we know that FMR has high heritability.\(^{40}\) Finally, given that more dispersive butterflies increase the colonization rate and decrease the extinction rate (rescue effect) (Table 3), the balance between extinctions and colonizations is shifted towards higher metapopulation size (equation (1)). In summary, there are biologically consistent associations between the SNP pgi:c.331A > C and individual traits (flight metabolism, dispersal rate), population processes (colonizations, extinctions) and landscape-level metapopulation attributes (metapopulation size and persistence).

Apart from \(Pgi\), the gene \(G6pd\) shows substantial polymorphism and significant association with landscapes structure (pooled area of habitat in the network). Genetic variation in metabolic enzymes, which often affect signalling pathways and have other moonlighting roles,\(^{41,42}\) is surprisingly often associated with fitness-related traits,\(^41\) but very little is known about the actual molecular functions. In Drosophila, polymorphism in \(G6pd\) and other enzymes around the glucose-6-phosphate (G6P) branching point of the glycolytic pathway often deviate from neutrality, suggesting that they may have been subject to adaptive evolution.\(^{33,44}\)

Allelic variation in \(Pgi\) is associated with life-history traits in many insects and plants,\(^{45}\) in which comparable results on population dynamics and size could be expected to occur. In other species living in highly fragmented landscapes, other heritable traits influencing dispersal and colonization may be selected for and be coupled with population dynamics in the same way as pgi:c.331A > C is in the Glanville fritillary. Examples range from gene expression profiles\(^{46,37}\) to wing polymorphism in insects\(^48\) and behavioural traits in vertebrates.\(^49\) We suggest that species inhabiting patchy habitats often exhibit such eco-evolutionary dynamics,\(^{32,30}\) reciprocal interaction between microevolutionary and demographic dynamics. A yet open question is how commonly eco-evolutionary dynamics increase the persistence of populations.\(^{50}\) Our results provide a convincing evidence that \(f_{\text{disp}}\) strongly affects metapopulation dynamics and as metapopulations close to the extinction threshold are more persistent if they have high \(f_{\text{disp}}\). Selection thus compensates, to a limited extent, for the adverse consequences of habitat loss.

**Methods**

**Habitat patches and patch networks.** In the Åland islands, the Glanville fritillary inhabits dry meadows that have at least one of the two larval host plant species, the ribwort plantain (Plantago lanceolata) or the spiked speedwell (Veronica spicata).\(^{25,26}\) The habitat patches are small: the median size is 0.06 ha, and only 1% are greater than 2 ha. We have mapped the entire Åland Islands, an area of \(50 \times 70\) km, for the habitat patches during two periods, in 1993 and 1998–1999. The first survey yielded ca. 1,500 patches, while the total number after the second survey has been ca. 4,500 patches, including a large number of very small ones. The exact current number is 4,415 (September 2015). We census the caterpillars annually in the overwintering stage by visiting all the habitat patches in late summer with the help of 50–70 field assistants.\(^{57}\) Hence, the census results reflect the presence of local breeding populations. The census is possible because caterpillars overwinter in groups of ca. 100 of mostly full-sibs under a relatively conspicious silken web woven at the base of the host plant (Fig. 1f). The locations of the found groups are marked with GPS and data is stored into the database in the field. A sample of three living larvae is collected from each group (1995, 2002, 2007–2012) or a subset of groups (2013–2014) or a subset of groups (2013–2014) or a subset of groups (2013–2014) or a subset of groups (2013–2014) or a subset of groups (2013–2014) or a subset of groups (2013–2014)) and taken to a butterfly rearing facility for further studies. Because only a third of the patches were censused before 1998–1999, the number of presence/absence records for individual years and populations is less than 22 \(\times\) 4,415. Additionally, there is a small amount of missing data especially from isolated low-quality networks that have remained completely unoccupied during the entire study period.

The 4,415 patches were divided into 125 sub-networks using the software SPOMSIM (ref. 51). The construction of patch networks is based on geometric average linkage clustering using connectivity (see ‘Connectivity of habitat patches’ below) as the distance measure and selecting a level of clustering that produced networks within which individual patches can be easily reached by dispersing butterflies, while movements between networks would be uncommon.\(^{52}\) In the clustering, we used the same parameter values as Molilanen;\(^{51}\) \(z = 1.0, b = 0.5\) and \(q = 1.5\), where \(z\) is the parameter of the negative exponential dispersal kernel (see ‘Connectivity of habitat patches’ below) and the other parameters control connectivity of the network. The final clustering is found in reality than this calculation suggests, because the clusters of patches (networks) are often separated dispersal barriers, such as tracts of forest. These landscape effects are not taken into account in the clustering algorithm, based on physical distances only.

**Analysis of control survey data.** Previously the probability of not recording an existing population (non-detection) has been estimated to range from 0.1 (ref. 36) to 0.28 (ref. 27). In this study, we estimate the prevalence of not detecting a larval group and not detecting a patch as occupied (Supplementary Table 1) using 111 control visits from the years 2009, 2011 and 2015. We also study the relationship of non-detection to local population size (Supplementary Table 3). The control data were only collected in networks where at least some larval groups were found during the main surveys. Out of 200 occupied patches, the control visits found 29 populations that were not recorded during the main survey. We assume that the remaining 111 patches were truly unoccupied, and therefore they are excluded from the following analyses.
The statistical analysis of control survey data consists of three generalized linear models. In the first model (Supplementary Table 1) we estimate the probability of detecting a larval group with a binomial regression model with a log-link function. As the dependent variable we use the proportion of larval groups found during regular survey, and as explanatory variables log-transformed patch area (ha) and total number of larval groups. The first explanatory variable tests the dependence of detecting a larval group on patch area, and the second for the dependence of detecting a larval group on total number of larval groups. As larval groups are not randomly distributed within patches, we expect larval groups to be more easily found when they occur in larger numbers.

In the second model (Supplementary Table 2) we analyse the overall probability of not detecting a patch as occupied when it is occupied. The dependent variable of the logistic regression is the binary indicator of non-detection, with an occupied patch being detected during the main survey coded as 0 and not detected as 1. The model uses the log-link function. The explanatory variables are log-transformed patch area (ha) and total number of larval groups. We also estimate the percentage of host plant areas surrounded by low vegetation and where competitively superior grasses are less growing in low vegetation and where competitively superior grasses are less growing in low vegetation. We also estimate the percentage of desiccated host plants in the common, we estimate the percentage of host plant areas surrounded by low vegetation and where competitively superior grasses are less growing in low vegetation and where competitively superior grasses are less growing in low vegetation.

Connectivity of habitat patches. Connectivity of patch i is a proxy for the number of immigrants arriving at patch i during one generation (year). Connectivity $S_i$ is calculated as $S_i = \sum_j A_{ij} e^{-d_{ij}/A_{em}}$, where $A_i$ and $A_j$ are the areas of the source (j) and target (i) patches (in ha), $p_j$ is the incidence of occupancy of source patch j, $d_{ij}$ is the distance between patches i and j (in km), and $e^{-d_{ij}}$ is the negative exponential dispersal kernel with parameter $\alpha$ (refs 33, 54). The incidence of occupancy $p_j$ has the value of 1 for occupied and 0 for unoccupied patches, or a value between 0 and 1 if the average value of $p_j$ across several years is used. The exponents $im$ and $em$ scale the rates of immigration and emigration by patch area. We estimated the values of $im$ and $em$ ($im = 0.44; em = 0.22$) are within the range of previous results for the Glanville fritillary and the False heath fritillary, a closely related butterfly with similar ecology55. The estimated value for $\alpha = 0.93$ agrees with earlier studies based on mark-recapture data56. In the model based on equation (2) and linear models where $z$ and $em$ cannot be estimated independently (see ‘Modelling extinction and colonization events’ below), we assume $z = 1$ and $em = 0.2$ based on the results above.

Colonization and extinction rates. Colonization rate of patch i is given by $c_{s_i}$, where $c_{s_i}$ is the colonization rate parameter. Extinction rate is assumed to depend on the area $A_i$ and quality $Q_i$ of patch i as $e^{(-c_{s_i} A_{i} Q_{i})}$, where $c_{s}$ and $c_{x}$ are parameters estimated from the present data (Table 2). The construction of the habitat quality variable $Q_i$ is described in the section ‘Parameter estimation based on equation (2)’ below.

Connectivity of patch networks. The nearest patch in other networks is located 1.75 km on average from the centre point of the focal network (minimum and maximum 0.62 and 5.13 km, respectively, $n = 125$), while the average distance to the ten nearest other patches is 2.17 km (0.95 and 6.12 km, respectively). Though most butterflies move < 1 km in their life-time, a substantial fraction, on the order of 10%, fly a distance of 2 km or more31, 65, 67, and hence we can expect some connectivity between the networks. We calculated a measure of connectivity for each network as $S_n = \sum_i e^{-d_{in}/N_i}$, where $N_i$ is the average population size (number of larval groups) in patch j across the years, and $d_{in}$ is the distance in km between the centre point of patch j and the centre point of network n. The sum is taken over all patches j that do not belong to network n. We assume $z = 1$ as in the calculation of connectivity for individual habitat patches.

Metapopulation capacity. Based on the assumptions of how patch areas, qualities and spatial locations in the network influence the extinction and colonization rates, one may construct an n by n matrix M for a network with n patches33, 54. The leading eigenvalue of M is called the metapopulation capacity and denoted by $\lambda_M$. In the present case, the elements of M are $m_{ij} = 0$ and $m_{ij} = A_{ij} c A_{em} e^{-d_{ij}/A_{em}}$, where $A_{ij} c$ is proportional to the expected life-time of population i (the inverse of extinction rate; see ‘Colonization and extinction rates’ above), while the remaining terms in $m_{ij}$ come from the assumptions of how connectivity depends on patch areas and configuration (see ‘Connectivity of habitat patches’ above).

Metapopulation capacity integrates the effects of patch areas, qualities and their spatial locations on the capacity of the network to support a viable metapopulation28, 33.

Parameter estimation based on equation (2). We estimated model parameters with non-linear regression using the expression for the probability of patch occupancy at equilibrium33, 54. Parameter estimation was carried out in two stages. In the first stage, we used equation where $\delta = c/c_{s}$ is the ratio of the extinction and colonization rate parameters, called the extinction threshold, and $x$ is the sum of the exponents $e\alpha$ and $im$ (see definition of $m_{ij}$ in the section ‘Metapopulation capacity’ above). $\hat{p}_i$ is the observed frequency of occupancy of patch i during the 22 years. Equation (3) was fitted to data from large networks with > 50 patches and average $\hat{p}_i > 0.05$, which are likely to have viable metapopulations. There are 24 such networks with 2,330 habitat patches. The estimated parameter values are $\delta = 3.91$ and $x = 0.51$.

We next used equation (3) to calculate the predicted values of $\hat{p}_i$ for each patch in the 24 large networks. Using logistic regression, we then explained the observed values of $\hat{p}_i$ with the predicted values as well as with four attributes of patch quality. Given that the effects of the four quality variables are of the same order of magnitude (Table 1), we summarized their effects with the term $y^5$, where $y$ is a parameter and $Q_i = q_{i0} + q_{i1} y + q_{i2} y^2 + q_{i3} y^3$. The sum of the four quality variables for patch i, all rescaled to the interval [−1, 1] (the sign of $q_{i0}$ is reversed because large values correspond to low quality; Table 1). Adding habitat quality with this term into the model, equation (3) is turned to equation (2) in the Results, which is repeated here.

In the second step, we fitted equation (2) to the data from the 24 large networks, and obtained the parameter values $\delta = 5.47$ (95% confidence interval 5.00–5.93), $x = 0.428$ (0.395–0.462) and $y = 1.71$ (1.63–1.79). Equation (2) with these parameter values predicts the observed patch occupancy as well as the logistic model with all the four patch quality variables added separately (Table 1). Therefore, we conclude that adding the term $y^5$ into the model is a simple and effective way of taking several features of habitat quality into account in the context of the present model.

Parameter estimation from colonization and extinction events. We also estimated model parameters using data on annual extinction and colonization events in all networks using the data for the years 1999–2014 (Table 2). The discrete-time extinction and colonization probabilities are given by $1 - \exp(-A_{ij} / c_{x})$ and $1 - \exp(-c_{s} A_{ij} Q_{ij})$, respectively (see ‘Colonization and extinction rates’ and ‘Connectivity of habitat patches’ above). The model was implemented using Stan version 2.8, and the parameters were estimated using Hamiltonian Markov Chain Monte Carlo57, 68. All parameters had uniformly distributed priors on the positive real numbers. The model was run with four chains for 1,000 warm-up steps and 1,000 steps of sampling. For all parameters the number of effective samples was >1,000, the convergence measure $R$ was <1.004 and the Monte Carlo standard error of the parameter means was <0.002.

Modelling extinction and colonization events. Colonizations and extinctions were analysed using a varying intercepts hierarchical logistic regression model with habitat patch and annual network-level random effects (Table 3). We used data for the years 1999–2014 in the analysis, because there is much missing data for the earlier years (see ‘Habitat patches and patch networks’ above). We also repeated the analysis for the years 1993–2014 and 2007–2014 with very similar results and the same conclusions (genetic data were collected in 2007–2012; see ‘Genetic data and measures’ below).

The patch-specific colonization and extinction events were estimated using patch area $A_{ij}^{0.2}$, connectivity $S_{ij}$ and $p_{ij}^{\text{obs}}$ as predictors. The structure of the
regression model is given by

\[ p_{it} \sim \text{Bernoulli}(\logit^{-1}(u_{it})) \]

\[ u_{it} = \theta_0 + \theta_1 + \theta_2 + \beta_1 S + \beta_2 X_{disp} \]

\[ x_0 \sim \text{Uniform}(-\infty, +\infty) \]

\[ x_2 \sim \text{Normal}(0, \sigma) \]

\[ x_3 \sim \text{Normal}(x_2, \sigma_2) \]

\[ \sigma_1 \sim \text{Cauchy}(0, \tau) \]

\[ \sigma_2 \sim \text{Cauchy}(0, \tau) \]

\[ \beta_1, \beta_2, \beta_3 \sim \text{Cauchy}(0, 5) \]

\[ p_{it} \sim \text{Gamma}(2, 0.1) \]

\[ f_{disp} \]

where \( p_{it} \) indicates the colonization or extinction event in patch \( i \) in year \( t \), \( x_0 \) is the mean immigration rate or colonization rate, \( x_2 \) is the random intercept for patch \( i \) in network \( n \). The annual network effects account for network structure and the variation in the scale of spatial synchrony between years. The network-specific patch effect accounts for differences between networks and repeated sampling of patches. \( f_{disp} \), \( f_{disp} \), and \( \beta_3 \) are the coefficients of spatial connectivity, annual sensitivity of dispersal, immigration, and patch area. Patch area (in ha) is scaled to power 0.2. The exponent 0.2 scales expected population size and the number of immigrants and emigrants by patch area, which approximates the parameters \( c, \text{im} \), and \( \text{em} \) (see ‘Connectivity of habitat patches’ above). The predictors \( A^2 \), \( S \), and \( f_{disp} \) are centred to zero mean to adjust for their relation to assist the comparison of linear regression model in R 3.3 (ref. 72). The sensitivity of linear regression analyses to other deviations from model assumptions, such as outliers, was tested by comparing coefficient estimates of the linear regression model to those from a robust linear regression model estimated with a Huber M-estimator using the R package MASS (ref. 74). None of the fitted models showed violations to heteroscedasticity or other model assumptions.

\[ x_1 \sim \text{Normal}(0, \sigma) \]

\[ x_2 \sim \text{Normal}(x_1, \sigma_2) \]

\[ \sigma_1 \sim \text{Cauchy}(0, \tau) \]

\[ \sigma_2 \sim \text{Cauchy}(0, \tau) \]

\[ f_{disp} \]

\[ p_{it} \sim \text{Gamma}(2, 0.1) \]

\[ f_{disp} \]

where \( p_{it} \) indicates the colonization or extinction event in patch \( i \) in year \( t \), \( x_0 \) is the mean immigration rate or colonization rate, \( x_2 \) is the random intercept for patch \( i \) in network \( n \). The annual network effects account for network structure and the variation in the scale of spatial synchrony between years. The network-specific patch effect accounts for differences between networks and repeated sampling of patches. \( f_{disp} \), \( f_{disp} \), and \( \beta_3 \) are the coefficients of spatial connectivity, annual sensitivity of dispersal, immigration, and patch area. Patch area (in ha) is scaled to power 0.2. The exponent 0.2 scales expected population size and the number of immigrants and emigrants by patch area, which approximates the parameters \( c, \text{im} \), and \( \text{em} \) (see ‘Connectivity of habitat patches’ above). The predictors \( A^2 \), \( S \), and \( f_{disp} \) are centred to zero mean to adjust for their relation to assist the comparison of linear regression model in R 3.3 (ref. 72). The sensitivity of linear regression analyses to other deviations from model assumptions, such as outliers, was tested by comparing coefficient estimates of the linear regression model to those from a robust linear regression model estimated with a Huber M-estimator using the R package MASS (ref. 74). None of the fitted models showed violations to heteroscedasticity or other model assumptions.

\[ x_1 \sim \text{Normal}(0, \sigma) \]

\[ x_2 \sim \text{Normal}(x_1, \sigma_2) \]

\[ \sigma_1 \sim \text{Cauchy}(0, \tau) \]

\[ \sigma_2 \sim \text{Cauchy}(0, \tau) \]

\[ f_{disp} \]

\[ p_{it} \sim \text{Gamma}(2, 0.1) \]

\[ f_{disp} \]

where \( p_{it} \) indicates the colonization or extinction event in patch \( i \) in year \( t \), \( x_0 \) is the mean immigration rate or colonization rate, \( x_2 \) is the random intercept for patch \( i \) in network \( n \). The annual network effects account for network structure and the variation in the scale of spatial synchrony between years. The network-specific patch effect accounts for differences between networks and repeated sampling of patches. \( f_{disp} \), \( f_{disp} \), and \( \beta_3 \) are the coefficients of spatial connectivity, annual sensitivity of dispersal, immigration, and patch area. Patch area (in ha) is scaled to power 0.2. The exponent 0.2 scales expected population size and the number of immigrants and emigrants by patch area, which approximates the parameters \( c, \text{im} \), and \( \text{em} \) (see ‘Connectivity of habitat patches’ above). The predictors \( A^2 \), \( S \), and \( f_{disp} \) are centred to zero mean to adjust for their relation to assist the comparison of linear regression model in R 3.3 (ref. 72). The sensitivity of linear regression analyses to other deviations from model assumptions, such as outliers, was tested by comparing coefficient estimates of the linear regression model to those from a robust linear regression model estimated with a Huber M-estimator using the R package MASS (ref. 74). None of the fitted models showed violations to heteroscedasticity or other model assumptions.

**Data availability.** Genetic variant data that support the findings of this study have been deposited in dbSNP with the accession codes ss2137343739 - ss2137343816. All other data that support the findings of this study are available from the corresponding author on reasonable request.

**References**

1. Millennium Ecosystem Assessment. Ecosystems and Human Well-being: Synthesis (Island Press, 2005).
2. Brooks, T. M. et al. Habitat loss and extinction in the hotspots of biodiversity. *Conserv. Biol. 16*, 909–923 (2002).
3. Pimm, S. L. et al. The biodiversity of species and their rates of extinction, distribution, and protection. *Science 344*, 987 (2014).
4. Pereira, H. M. et al. Scenarios for global biodiversity in the 21st century. *Science 330*, 1496–1501 (2010).
5. Keith, D. A. et al. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Bioll. Lett. 4*, 560–563 (2008).
6. Fahrig, L. Rethinking patch size and isolation effects: the habitat amount hypothesis. *J. Biogeogr. 40*, 1649–1663 (2013).
7. Hanski, I. Habitat fragmentation and species richness. *J. Biogeogr. 42*, 989–993 (2015).
8. Didham, R. K., Kapos, V. & Ewers, R. M. Rethinking the conceptual foundations of habitat fragmentation research. *Oikos 121*, 161–170 (2012).
9. Ruffell, J., Banks-Leite, C. & Didham, R. K. Accounting for the causal basis of collinearity when measuring the effects of habitat loss versus habitat fragmentation. *Oikos 125*, 117–125 (2016).
10. Rosenzweig, M. L. Species diversity in space and time (Cambridge University Press, 1995).
11. Pimm, S. L. & Askins, R. A. Forest losses predict bird extinctions in eastern North America. *Proc. Natl Acad. Sci. USA 92*, 9343–9347 (1995).
12. May, R. M., Lawton, J. H. & Stork, N. E. in Extinction Rates (eds Lawton, J. H. & May, R. M.) 1–24 (Oxford University Press, 1995).
13. Brooks, T. & Balmford, A. Atlantic forest extinctions. *Nature 380*, 115 (1996).
14. Pimm, S. L. & Raven, P. Biodiversity–extinction by numbers. *Nature 403*, 843–845 (2000).
15. He, F. L. & Hubbell, S. P. Species–area relationships always overestimate extinction rates from habitat loss. *Nature 473*, 368–371 (2011).
16. Kinzig, A. P. & Harte, J. Implications of endemics–area relationships for estimates of species extinctions. *Ecology 81*, 3305–3311 (2000).
17. Rybicki, J. & Hanski, I. Species–area relationships and extinctions caused by habitat loss and fragmentation. *Ecol. Lett. 16*, 27–38 (2013).
18. Hanski, I., Zurita, G. A., Bellocq, M. I. & Rybicki, J. Species–fragmented area relationship. *Proc. Natl Acad. Sci. USA 110*, 12715–12720 (2013).
19. Debinski, D. M. & Holt, R. D. A survey and overview of habitat fragmentation experiments. *Conserv. Biol. 14*, 342–355 (2000).
20. Hubbell, S. P. The unified neutral theory of biodiversity and biogeography (Princeton University Press, 2001).
21. Hallam, A. & Wignall, P. B. Mass extinctions and sea-level changes. *Earth-Sci. Rev. 48*, 217–250 (1999).
22. Hanski, I. The shrinking world: ecological consequences of habitat loss (International Ecology Institute, 2005).
23. Tscharnkte, T. et al. Landscape moderation of biodiversity patterns and processes—eight hypotheses. *Bioll. Rev. 87*, 661–685 (2012).
24. Ribeiro, M. C., Metzger, J. P., Martensen, A. C., Ponzonni, F. J. & Hirota, M. M. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. *Biol. Conserv.* **142**, 1141–1153 (2009).

25. Hanski, I. Metapopulation ecology (Oxford University Press, 1999).

26. Hanski, I. Eco-evolutionary spatial dynamics in the Glanville fritillary butterfly. *Proc. Natl Acad. Sci. USA* **108**, 14397–14404 (2011).

27. Östen, S., Nieminen, M., Meyke, E., Pöyry, J. & Hanski, I. Long-term metapopulation study of the Glanville fritillary butterfly (*Melitaea cinxia*): Survey methods, data management, and long-term population trends. *Ecol. Evol.* **3**, 3713–3737 (2013).

28. Hanski, I. & Ovaskainen, O. The metapopulation capacity of a fragmented landscape. *Nature* **404**, 755–758 (2000).

29. Carroll, S. P., Hendry, A. P., Reznick, D. N. & Fox, C. W. Evolution on ecological time-scales. *Func. Ecol.* **21**, 387–393 (2007).

30. Pelletier, F., Garant, D. & Hendry, A. P. Eco-evolutionary dynamics. *Introduction*. *Phil. Trans. R. Soc. B* **364**, 1483–1489 (2009).

31. Hanski, I. in *Dispersal ecology and evolution* (eds Clobert, J., Baguette, M., Benton, T. G., Bullock, J. M.) 290–303 (Oxford University Press, 2012).

32. Ovaskainen, O. & Hanski, I. How much does an individual habitat fragment contribute to metapopulation dynamics and persistence? *Theor. Popul. Biol.* **64**, 481–495 (2003).

33. Ovaskainen, O. & Hanski, I. in *Ecology, Genetics, and Evolution in Metapopulations* (eds Hanski, I., Gaggiotti, O. E.) 73–103 (Elsevier Academic Press, 2004).

34. Niitopöld, K. et al. Flight metabolic rate and Pgi genotype influence butterfly dispersal rate in the field. *Ecology* **90**, 2223–2232 (2009).

35. Hanski, I. & Mononen, T. Eco-evolutionary dynamics of dispersal in spatially heterogeneous environments. *Ecol. Lett.* **19**, 1025–1034 (2011).

36. Nieminen, M., Siljander, M. & Hanski, I. in *On the wings of Checkerspots: a model system for population biology* (eds Ehrlich, P. R. & Hanski, I.) 63–91 (Oxford University Press, 2004).

37. Tack, A. J. M., Mononen, T. & Hanski, I. Increasing frequency of low summer precipitation synchronizes dynamics and compromises metapopulation stability in the Glanville fritillary butterfly. *Proc. R. Soc. Lond. B Biol. Sci.* **282**, 20150173 (2015).

38. André, H. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *Oikos* **71**, 355–366 (1994).

39. Niitopöld, K. Genotype by temperature interactions in the metabolic rate of the Glanville fritillary butterfly. *J. Exp. Biol.* **213**, 1042–1048 (2010).

40. Mattila, A. L. K. & Hanski, I. Heritability of flight and resting metabolic rate in the Glanville fritillary butterfly. *J. Evol. Biol.* **27**, 1733–1743 (2014).

41. Marden, J. H. Nature’s inordinate fondness for metabolic enzymes: why metabolic enzyme loci are so frequently targets of selection. *Mol. Ecol.* **22**, 5743–5764 (2013).

42. Kim, J. W. & Dang, C. V. Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* **30**, 142–150 (2005).

43. Flowers, J. M. & Dang, C. V. Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* **30**, 142–150 (2005).

44. Niitopöld, K. Genotype by temperature interactions in the metabolic rate of the Glanville fritillary butterfly. *J. Exp. Biol.* **213**, 1042–1048 (2010).

45. Mattila, A. L. K. & Hanski, I. Heritability of flight and resting metabolic rate in the Glanville fritillary butterfly. *J. Evol. Biol.* **27**, 1733–1743 (2014).

46. Marden, J. H. Nature’s inordinate fondness for metabolic enzymes: why metabolic enzyme loci are so frequently targets of selection. *Mol. Ecol.* **22**, 5743–5764 (2013).

47. Kim, J. W. & Dang, C. V. Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* **30**, 142–150 (2005).

48. Flowers, J. M. & Dang, C. V. Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* **30**, 142–150 (2005).

49. Duckworth, R. A. & Krouk, L. E. B. Evolution of genetic integration between dispersal and colonization ability in a bird. *Evolution* **63**, 968–977 (2009).

50. Kinnison, M. T. & Hairston, N. G. Eco-evolutionary conservation biology: contemporary evolution and the dynamics of persistence. *Func. Ecol.* **21**, 444–454 (2007).

51. Moilanen, A. SPOCSIM: software for stochastic patch occupancy models of metapopulation dynamics. *Ecol. Model.* **179**, 533–550 (2004).

52. Hanski, I., Moilanen, A., Pakkala, T. & Kuussaari, M. The quantitative incidence function model and persistence of an endangered butterfly metapopulation. *Conserv. Biol.* **10**, 578–590 (1996).

53. Gabry, J. & Goodrich, B. rstanarm: Bayesian Applied Regression. Modelling via Stan. Version 2.9. http://mc-stan.org/interfaces/rstanarm.html (2016).

54. Hanski, I. A practical model of metapopulation dynamics. *J. Anim. Ecol.* **63**, 151–162 (1994).
Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Hanski, I. et al. Ecological and genetic basis of population persistence of the Glanville fritillary butterfly in fragmented landscapes. Nat. Commun. 8, 14504 doi: 10.1038/ncomms14504 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2017