SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
BEHAVIOURAL ECOLOGY
ECOLOGICAL GENETICS
EVOLUTIONARY GENETICS
EVOLUTIONARY THEORY

Received
5 April 2013

Accepted
6 January 2014

Published
5 February 2014

Correspondence and requests for materials should be addressed to E.D. (etienne.danchin@univ-tlse3.fr)

Avoiding pitfalls in estimating heritability with the common options approach

Etienne Danchin[1,2], Eric Wajnberg[3] & Richard H. Wagner[4]

[1]CNRS, UPS, ENFA; EDB (Laboratoire Evolution & Diversité Biologique); UMR5174; 118 route de Narbonne, F-31062 Toulouse, France, [2]Université de Toulouse; UMR5174; F-31062 Toulouse, France, [3]INRA, 400 Route des Chappes, BP 167, 06903 Sophia Antipolis Cedex, France, [4]Konrad Lorenz Institute of Ethology, Department of Comparative Biology and Evolution, University of Veterinary Medicine, Vienna, Savoyenstrasse 1a, 1160 Vienna, Austria.

In many circumstances, heritability estimates are subject to two potentially interacting pitfalls: the spatial and the regression to the mean (RTM) fallacies. The spatial fallacy occurs when the set of potential movement options differs among individuals according to where individuals depart. The RTM fallacy occurs when extreme measurements are followed by measurements that are closer to the mean. We simulated data from the largest published heritability study of a behavioural trait, colony size choice, to examine the operation of the two fallacies. We found that spurious heritabilities are generated under a wide range of conditions both in experimental and correlative estimates of heritability. Classically designed cross-foster experiments can actually increase the frequency of spurious heritabilities. Simulations showed that experiments providing all individuals with the identical set of options, such as by fostering all offspring in the same breeding location, are immune to the two pitfalls.

Understanding the evolution of ecological adaptations requires the measurement of heritability, which is the proportion of phenotypic variation that is genetically transmitted to offspring[1–4]. Heritability has long been identified in numerous life history, physiological and morphological traits[1,4]. It has also been reported in such complex behavioural traits as dominance, aggression[5], dispersal[6], personality[7–12], cooperative breeding[13] and group size choice[14]. These findings are striking because measuring heritability of behavioural traits, especially in the field, is a daunting task[3,4]. A reason for this is that the constraints on individuals to make optimal choices create considerably more variation than in other characteristics such as morphological traits. In this context, Brown and Brown[14] reported exceptionally high heritabilities of individual preferences for colonies of particular sizes. Animals in many species forage, travel or breed in groups, and group size often varies by orders of magnitude across species and populations[15–17]. Studying variation in group size therefore, is a widely applied approach for understanding group living[16–20]. A taxonomically widespread form of group living is omit coloniality, in which breeders defend only relatively small, aggregated breeding territories and forage elsewhere[16,21].

A novel solution for explaining variation in colony sizes has been offered by Brown and Brown[14] based on their study of cliff swallows *Petrochelidon pyrrhonota*, whose colonies range from two to over 3000 nesting pairs[18]. Brown and Brown[14] proposed that variation in colony size is maintained by a genetic predisposition by breeders to recruit to colonies of similar sizes to those chosen by their parents. This idea was supported by highly significant parent-offspring regressions of colony size ranks, suggesting that colony size choice is heritable. To exclude the possibility that these findings could be explained by non-genetic factors such as early social imprinting, Brown and Brown[14] performed a partial cross-foster experiment in which half of the nestlings in broods in small colonies were transferred to be raised in large colonies, and *vice versa*. The authors found that their populations showed significant positive regressions to the natal colonies and negative regressions to the colonies in which individuals were raised. Similar results have been subsequently reported in an experimental study of barn swallows *Hirundo rustica*[22] and in a correlational study of lesser kestrels *Falco naumanni*[23].

Genetic transmission of a complex behavioural trait from parents to offspring may have considerable evolutionary consequences. The cliff swallow study is especially compelling given its strong results and the extraordinarily large sample sizes, which were produced by the cross-fostering of almost 2000 nestlings of which 721 were recovered as breeders in the following year[14]. For these reasons the study has been hailed as a milestone[24].

These impressive results are however unexpected for at least three reasons. First, compared to morphological traits, the high plasticity of realized behaviour makes behavioural traits unlikely to be highly heritable[4]. Second, habitat selection is strongly influenced by the spatial distribution of available breeding locations. Previous studies

have stressed that heritability estimates of behavioural traits that involve movement by animals across distances can be strongly inflated when not considering that different individuals have different sets of possible movements[25,26]. The sets of possible outcomes are affected by such factors as the distribution of suitable breeding sites, the locations of natal nests and the shape of the study areas. For example, animals born in the centre of a study area have a different set of possible dispersal distances than those born in the periphery. Because dispersal distance can only be studied in individuals that remain within the study area, this method produces a bias toward individuals with the same short dispersal distances as their parents. If offspring also disperse relatively short distances, as do most individuals in many species[6,26,27], it could produce spurious parent-offspring regressions. This spatial fallacy can be circumvented by using a null model of possible choices accounting for the spatial distribution of potential movements for every individual[25]. However, none of the three studies reporting on heritability of colony size choice used such a null model of possible choices, suggesting that heritabilities may have been overestimated.

Third, parent-offspring regressions are well known to be particularly prone to the ubiquitous fallacy called "regression to the mean" (RTM) that was first identified in the 19th century[28] in the context of parent-offspring regressions. The RTM fallacy results from the fact that uncommonly large or small measurements are generally followed by measurements that are statistically closer to the mean simply because average values are far more common than extreme ones. In cross-foster experiments that study the heritability of colony size choice, when individuals are fostered from small to large colonies they will on average recruit to colonies that are statistically smaller than their foster colony because these recruitment colonies are closer to the mean colony size (and *vice versa*).

The ecological and evolutionary implications of a genetic component of group size choice are profound. We have thus explored, using simulations of published data, the potential of the RTM and spatial fallacies to produce spurious parent-offspring regressions in the context of colony size choice. Because, to our knowledge, the impact of the RTM and spatial fallacies has never been explored in tandem, we also examined the interaction between the potential effects of the two fallacies on estimated heritabilities. We finally applied this approach to develop a method for avoiding both pitfalls.

Results

Simulating heritabilities from an experimental study. We used individual based simulations to explore the potential occurrence of the two fallacies and their impact on estimated heritabilities (See Methods). Our simulations of the cliff swallow experiment[14] randomly produced a high proportion of regressions on recruitment colony size that were positive to birth and negative to foster colony size (Figure 1a). This finding was obtained at both significance thresholds and in all four spatial colony distributions. The first three distributions simulated the natural condition of large colonies being widely spaced and surrounded by smaller colonies. Of these, the BigFar5 distribution was designed to generate the maximum contrast in the sizes of neighbouring colonies. That distribution produced regressions with equal or lower p-values than those of the cliff swallow study in 91% of the simulations (black bars in Figure 1a). The other two distributions in which large colonies were widely spaced, BigFar5Random and BigFarHalf, both randomly yielded over 50% of regressions with equal or lower p-values than in the cliff swallow study. The Random distribution of colony sizes comprised a null model in that it was generated without any assumptions about the spatial distributions of colonies, yet even it yielded equal or lower p-values in 33% of the simulations. When using the significance threshold of 0.05, the proportion of significant regressions increased only slightly in all four distributions (white bars in Figure 1a) because the distributions of the generated p-values were strongly skewed toward highly significant regressions. In the data in Figure 1a, for instance, between 35% and 72% of the p-values were lower than 0.0001. These findings suggest that highly significant heritabilities should be viewed with caution.

All of these results were generated with the random walk process of recruitment, which does not involve active choices of colonies but accounts for their spatial distributions. This recruitment strategy reproduces the fact that the recruitment probability rapidly declines with distance to the natal site[6,26,27]. These findings highlight the importance of accounting for the spatial distributions of potential choices and suggest that even in experiments, it is impossible to properly estimate heritabilities from parent-offspring regressions in the absence of a null model.

Simulating heritabilities when birth and foster colonies are randomly chosen. We then conducted simulations to analyze factors that may have contributed to the significant positive regressions to the birth colony and negative regressions to the foster colony in the cliff swallow study. Simulating an experimental design that randomly selects birth and foster colonies substantially diminished the percentage of significant regressions (compare Figures 1a and 1b). However, the percentage of significant parent-offspring regressions still ranged from 14% to 50% versus the expected 2.5%, indicating that randomization does not solve the problem.

Simulating non-experimental heritabilities. The three studies of the heritability of colony size choice[14,22,23] reported highly significant parent-offspring regressions with non-experimental correlations. Our simulations of data from Brown and Brown's Table 1 showed that significant parent-offspring regressions of the predicted signs were much less likely to be randomly generated from the non-experimental than the experimental data in three of our four distributions (compare Figure 1a and 1c). These were the three distributions with spatially structured colonies. In contrast, when colonies of various sizes were randomly distributed, the percentage of significant regressions of the expected sign was similar in the experimental and non-experimental data. This exercise illustrates how spatial structuring can generate spurious significant parent-offspring regressions. These differences exist even though our random colony distribution is conservative in that it generates some proportion of spatial structuring. The highly spatially structured BigFar5 colony distribution illustrates this point as it was designed to generate the maximum contrast in the sizes of neighbouring colonies to depict the highest risks of producing spurious experimental regressions. The consequence is that in non-experimental data, this distribution generated no significant regressions of the predicted signs (Figure 1c) but 99% of the opposite signs. This occurred because large colonies were surrounded by small ones, making it inevitable that most birds that fledged from large colonies (whether birth or foster colonies) would recruit to small ones.

The other two spatially structured distributions generated a substantial proportion of significant non-experimental parent-offspring regressions of the predicted signs (Figure 1c). There was thus substantial overlap among the distributions in their capacity to produce spurious significant experimental and non-experimental regressions of the predicted signs (Figure 1). This suggests that regardless of the type of colony distribution, spurious regressions occur. Thus, both experimental (Figure 1a) and non-experimental (Figure 1c) regressions can be highly significant and yet spurious.

The spatial and RTM fallacies. We next explored possible mechanisms responsible for the frequently generated spurious regressions in the experimental data. We found that regressions from cross-foster data are sensitive to differences in the sizes of birth and foster colonies (Figure 2). When nestlings were exchanged between

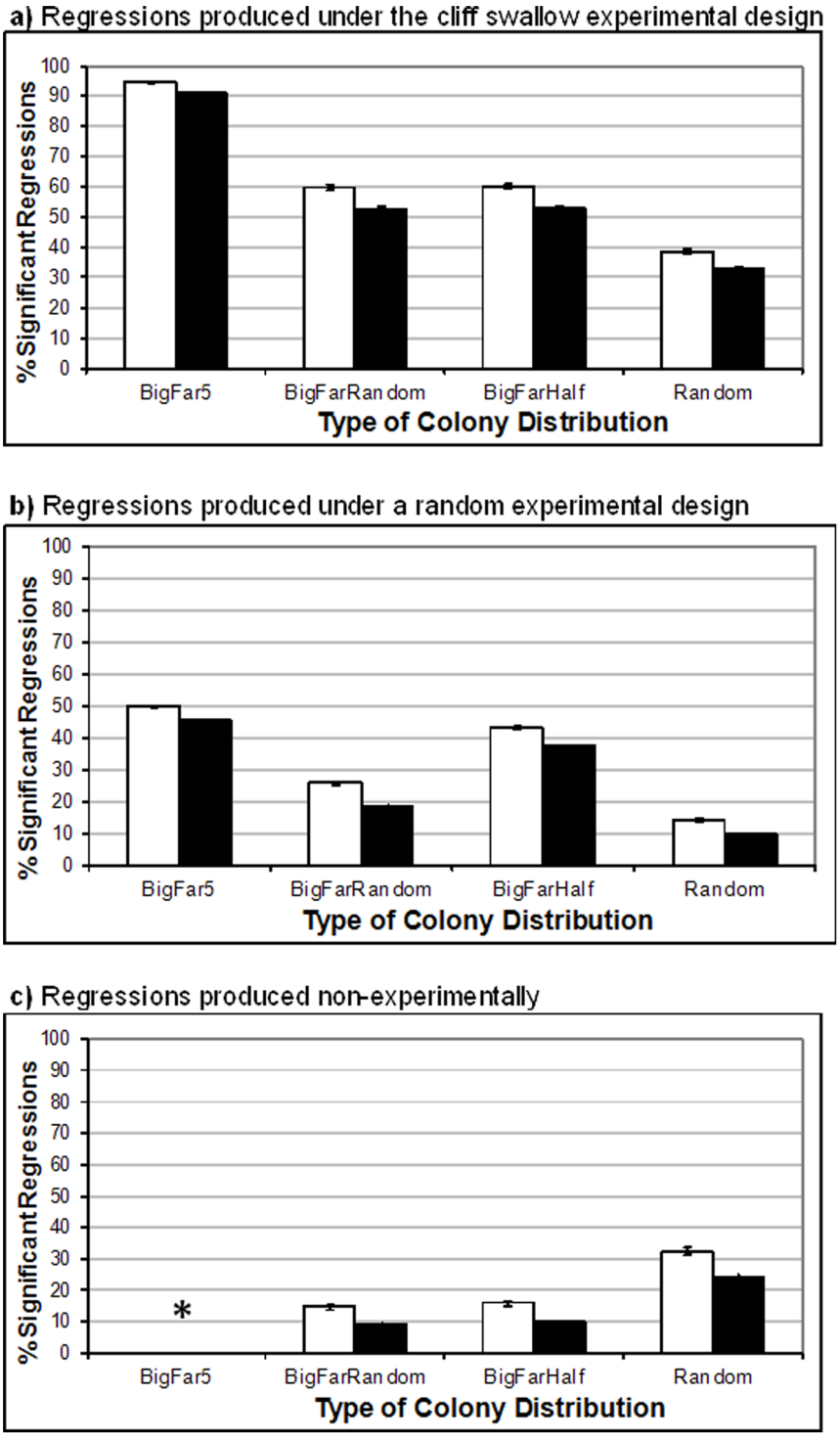

Figure 1 | Percentage (±SE) of randomly generated significant parent-offspring regressions in relation to four types of colony distributions. Individuals recruited from their foster (a and b) or birth (c) colony using a random process within a linear lattice. Open bars: when using 0.05 as the significance threshold; Black bars: when using the p-values of the cliff swallow regressions as the significance threshold. As explained in Methods, results only include instances when the parent-offspring regression was positive to the birth colony and negative to the foster colony. Each situation was simulated 2,000 times. (a) Simulations of the cross-foster experiments using the same protocol and sample sizes as in the cliff swallow study's Figure 2. (b) Simulations of the cross-foster experiments using the same sample sizes as in the cliff swallow study, but in which the birth and cross-foster colonies were selected randomly. (c) Simulations of the non-experimental parent-offspring regressions using the data provided in the cliff swallow study's Figure 1. Results only account for non-philopatric individuals. *: in this distribution there were no significant parent-offspring regressions of the expected signs, but 99% of the opposite signs.

Table 1 | Significance of the regressions shown in Figures 2 and 3. Regressions are of the percentage of significant parent-offspring heritabilities (Y-axis) on contrasts in size ranks between the birth and foster colonies (X-axis). All slopes were positive, suggesting that RTM was present in all situations. The greater the contrast in the sizes of natal and foster colonies, the more frequently spurious significant parent-offspring regressions were generated. There were 21 colonies in all simulations. All individuals were recruited using the random walk process in which recruits returned to the colony from which they fledged. They next moved randomly, recruiting to the first colony they encountered. When running simulations with 10 and 100 colonies we found similar results (Figure 3). In all circumstances the frequencies of significant parent-offspring regressions with low contrasts in the size of natal and birth colonies were from 4 to 8 times higher than the expected 2.5%

Lattice type	Linear lattice		Square lattice*	
Colony distribution	r^2	P	r^2	P
BigFar5	**0.70**	**0.0025**	**0.80**	**0.0005**
BigFar5Random	**0.82**	**0.0003**	**0.87**	**<0.0001**
BigFarHalf	**0.97**	**<0.0001**	**0.995**$^{\mu}$	**<0.0001**$^{\mu}$
Random	**0.61**	**0.0079**	**0.92**	**<0.0001**

*The four situations represented in Figure 2.
μ Situation of Figure 3.

large and small colonies (high contrast), more positive regressions were generated with the birth colony and more negative regressions with the foster colony than when they were exchanged between colonies of intermediate sizes (low contrast). The colony distributions were held constant within each of the curves in Figures 2 and 3. Thus, the increase in the proportions of significant regressions could only result from the occurrence of RTM.

The four colony distributions were each simulated in a linear and a square lattice. The percentage of significant heritabilities increased with the size contrast between birth and foster colonies (Figure 2). All slopes in Figure 2 were positive and significant, even when colony sizes were randomly distributed (statistics in Table 1). Similar results were obtained by running these simulations with 10, 20 and 100 colonies (Figure 3), suggesting that the frequency of spurious regressions was independent of the number of colonies in the experimental population.

When using the lowest possible size contrast between birth and foster colonies, RTM should be weak. Nevertheless, the proportion of significant regressions was approximately four to eight times higher than the expected significance threshold of 0.025 (left part of Figures 2 and 3). In the relative absence of RTM, these findings suggest the existence of another effect that produces spurious regressions. Figure 2 illustrates that the slopes produced in the random colony distribution are much lower than in the three spatially structured distributions. This difference suggests that the spatial fallacy is responsible for these spurious regressions. Results in Figures 2 and 3 thus suggest that the spatial fallacy and RTM combine to produce spurious regressions. The combined effects of the two pitfalls have unexpectedly strong consequences, with the frequency of spurious regressions of the expected signs ranging from 38% to 95% among the four colony distributions (Figure 1a).

The common options approach. We used further simulations to explore the conditions under which we would obtain the expected 2.5% of significant parent-offspring regressions in a one-tailed test. When we selected the median sized colony as the common foster colony, the simulations yielded 2.5% of significant regressions that

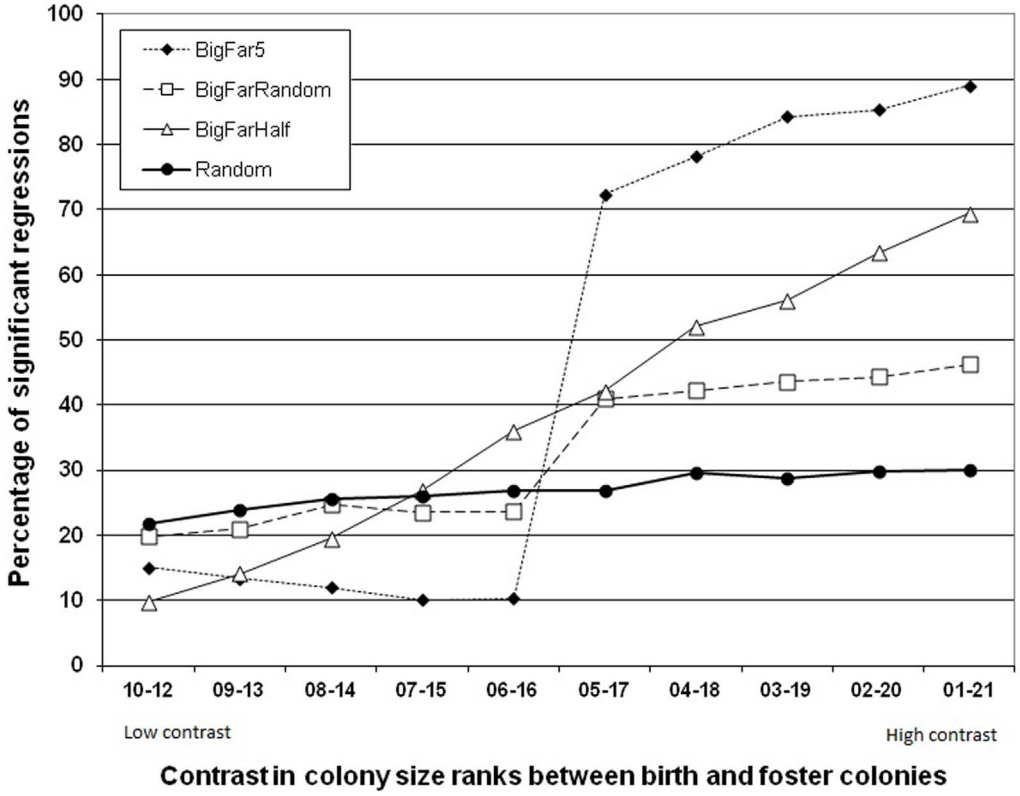

Figure 2 | The effect of contrasts in colony size ranks between birth and foster colonies on the percentage of significant parent-offspring regressions. We conservatively used the p-values from Table 1 of[14] as significance thresholds, and used the sample sizes from that study's Figure 2. There were 21 colonies and a square grid in all simulations (a linear grid leads to similar patterns; see Table 1). Recruitment was according to a random walk. Each point results from 2,000 simulations. The four regressions depicted here were significant (see Table 1). Standard errors are too small to be shown.

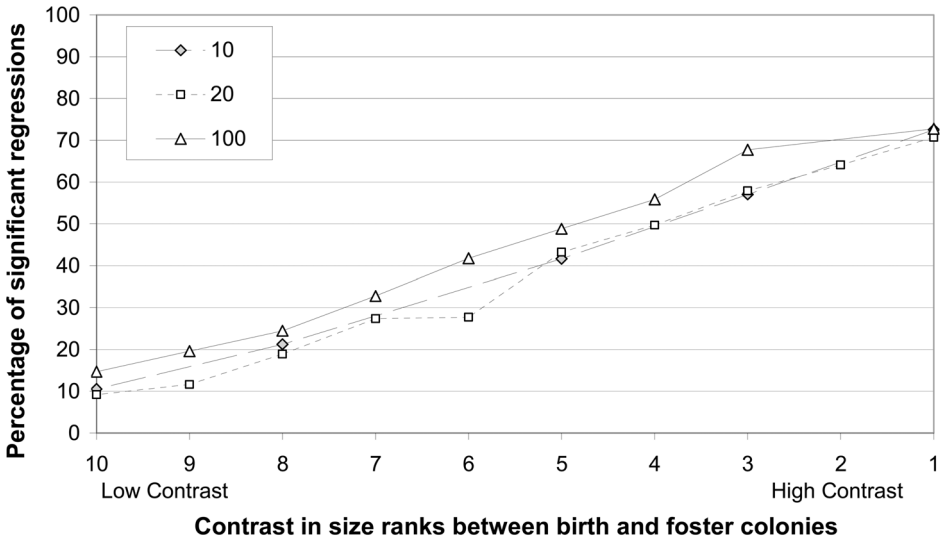

Figure 3 | Effect of the contrast in size ranks between birth and foster colonies on the percentage of significant parent-offspring regressions of colony sizes. Contrast in size ranks between birth and foster colonies ranged on the X-axis from the lowest (value of 10) to the highest (value of 1). For example, with 100 colonies, the lowest contrast was between colony sizes 50 and 51 while the highest contrast was between sizes 1 and 100. All three regressions depicted here were significant (SAS GLM procedure, interaction $p = 0.37$; number of colonies: $p = 0.028$; Effect of the design: $P < 0.0001$). We used the BigFarHalf distribution of colonies on a square grid and recruitment followed a random walk. Each point was produced from 2,000 simulations. Standard errors are too small to be shown.

were positive to the birth colony and negative to the foster colony (Figure 4a). This was true in all four types of colony distributions (Figure 4a), suggesting the robustness of this design. When we selected any colony size as the common foster colony, we obtained the same result (Figure 4b). The common options approach thus seems immune to the two pitfalls.

Discussion

Our simulations highlight two pitfalls in estimating the heritability of behavioural traits. One pitfall is regression to the mean (RTM), which was first identified in the context of parent-offspring regressions in the 19th century[28]. The other pitfall is the spatial fallacy, which may occur in studies of behavioural traits involving movements by animals between a set of potential locations.

RTM results from the fact that uncommonly large or small measurements are generally followed by measurements that are statistically closer to the mean simply because average values are far more common than extreme ones. This fallacy has been undermining studies in many fields such as economics, social sciences, cognition, health care policy, sports medicine and epidemiology despite its early discovery[28]. The current use of the term 'regression' is itself derived from the regression to the mean fallacy.

Our simulations show that analyses of data from cross-foster experiments may be sensitive to differences in the sizes of birth and foster colonies. This is potentially important because the greater the difference, the higher the probability that an individual will randomly recruit to a colony that is substantially different in size from the one in which it was fostered. For example, in a population with n colonies, if offspring are transferred from the largest to the smallest colony and then randomly select a colony, they have $(n - 1)/n$ probability to recruit to colonies that are larger than their foster colony. This may generate spurious significant positive correlations between the sizes of recruitment and birth colonies and negative ones between recruitment and foster colonies. Despite its ubiquity and venerability, RTM is so subtle and counter-intuitive that it continues to be overlooked[29]. It can be corrected for in post hoc analyses[30,31], but is considered to be avoidable by proper experimental designs[31]. Unfortunately, as our simulations suggest, even classically designed experiments are not immune to RTM and may even amplify its effects.

Accordingly, when our simulations varied the contrast in colony sizes, we found that the greater the contrast, the higher the proportion of spurious significant regressions (Figures 2 and 3). The contrasts in colony sizes may further explain why simulating randomly selected birth and foster colonies did not avoid RTM. In such a design, some of the randomly generated pairs of colonies inevitably have high contrasts in size and thus generate spurious regressions. Our exercises demonstrate the subtlety of RTM[29] and imply that randomization may not be an ubiquitous solution for avoiding biases in experimental designs.

Moreover, RTM mainly results from the bell shaped frequency distributions of most traits. This is due to the fact that the most frequently occurring values are of intermediate measures. In our context, the frequency distributions of colony-sizes are probably bell shaped because colonies of intermediate sizes are much more frequent than colonies of extreme sizes. In the cliff swallow study however, the frequency distributions of colony-size were flattened by the use of size ranks, with each rank being represented by a single colony. Thus, the effects of RTM became detectable only with increased contrasts in colony sizes. This explains the increase in the proportions of significant regressions in Figures 2 and 3. Consequently, the use of ranks in colony sizes in the cliff swallow study substantially diminished the effect of RTM, making our simulations conservative. This factor indicates that RTM may arise even in the absence of a bell shaped frequency distribution of the concerned trait, and highlights the importance of its impact in cross-foster experiments.

In the cliff swallow experimental design, chicks were cross-fostered between large and small colonies. This is intuitively appealing because experimenters seek to produce strong effects. However, our simulations show that spurious regressions can also be obtained even when chicks are swapped between colonies of low contrasted sizes (Figures 2 and 3). This suggests that an additional fallacy is involved.

The spatial fallacy was raised by van Noordwijk[25], who questioned whether a study that reported significant non-experimental parent-offspring regressions of dispersal distance in great tits *Parus major*[32] could conclude that dispersal distance is heritable. van Noordwijk[25] simulated data on between-nest box distances from three popula-

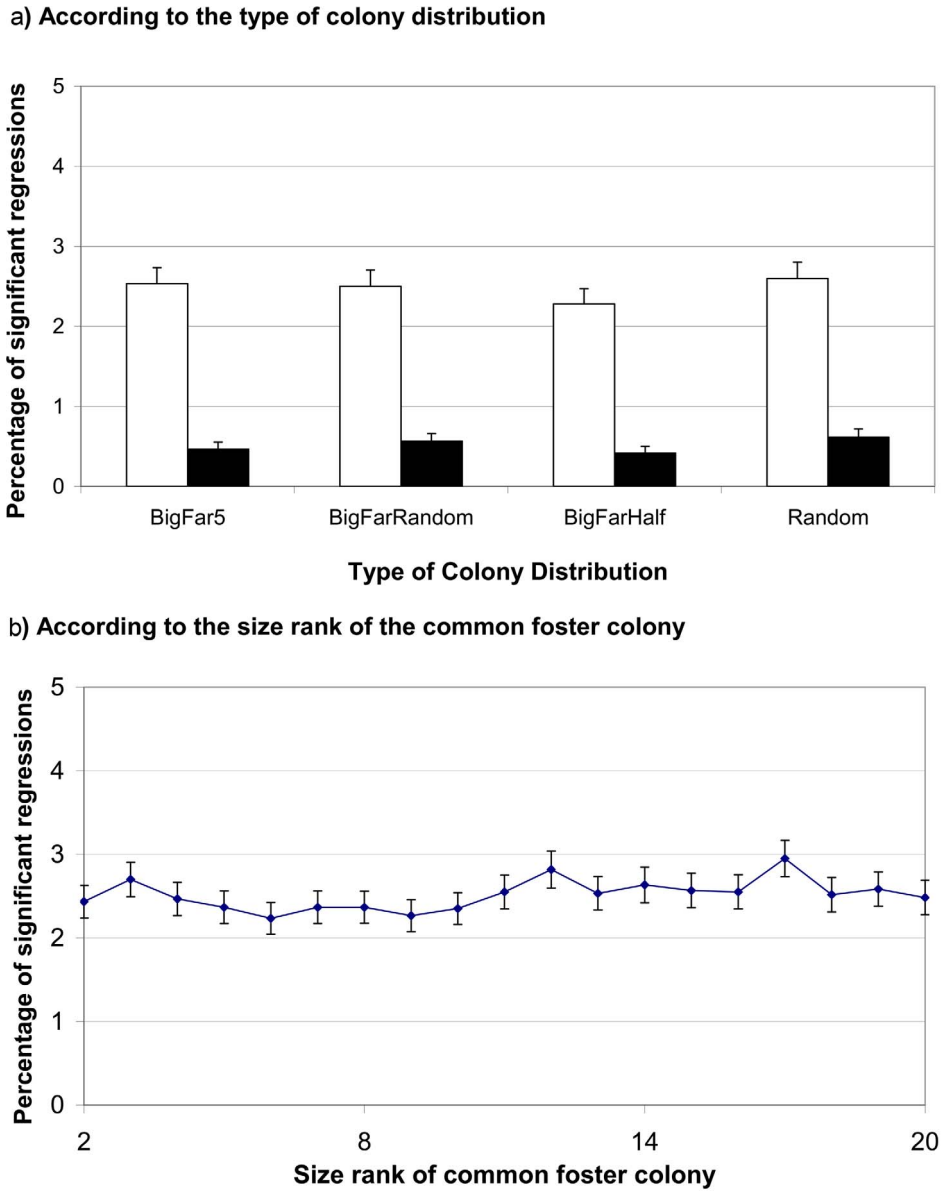

Figure 4 | The "Common Options" experimental design. Percentage (±SE) of significant parent-offspring regressions that were randomly generated according to four types of distributions of colonies. There were 21 colonies. Individuals recruited from the foster colony with a random walk. **Open bars**: when using 0.05 as the significance threshold; **Black bars**: when using the p-values of the cliff swallow regressions as the significance threshold. We only tabulated regressions that were positive to the birth, and negative to the foster colonies. Each situation was simulated 2,000 times. (Note that the scale of the Y-axes differs from those of Figures 1, 2 and 3.) (a) When transferring chicks from randomly selected colonies to the median colony (of size rank 11). The two parameters are: the BigFar5 colony distribution, and the transferring of chicks from the largest and smallest colony (1 and 21) into the median size colony (rank 11). (b) Analysis of the effect of the size of the common foster colony on the percentage of significant parent-offspring regressions ($P < 0.05$). The two parameters are: the BigFar5 colony distribution, and the transferring of chicks from the largest and smallest colony (1 and 21) into a single colony whose size was allowed to vary from ranks 2 to 20.

tions of great tits to generate sets of all possible inter-nest distances. His simulations led him to conclude that space must be accounted for in all correlative studies involving the distribution of suitable habitat in the environment. The simulations clearly showed that in a non-experimental study, spurious parent-offspring regressions of dispersal distances can be generated in the absence of a null model of all possible options. This method was recently applied in the context of dispersal in lesser kestrels[26], which found that philopatry to the natal colony was much higher, and observed distances much lower, than predicted by a null model accounting for all possible distances. More recently, van Noordwijk and collaborators designed methods to account for the impact of the heterogeneities in detectability of individuals that may result from differences in personality or sex on estimated heritabilities[33,34].

In the three studies of heritability of colony size choice[14,22,23], the distribution of colonies is comparable to the distribution of nest boxes in the great tit study in that there is also a finite set of choices of breeding locations. Consequently, estimating heritability of colony size choice requires taking into account the distribution of colonies of various sizes, which includes the number and density of colonies, the suitability of habitat, and inter-colony distances. For example, it is known that large colonies tend to be spaced relatively far apart, with smaller colonies situated in between[35–37] due to local competition for food[38]. Thus, as in dispersal studies[25,26], it is necessary to incorporate

into heritability estimates the randomly generated expected distributions of choices resulting from the spatial distribution of colonies of various sizes. By not doing so, the three studies assumed that individuals were equally likely to recruit to any of the colonies, regardless of their size and location. However, the probability of a given bird recruiting to a colony is likely to decrease with the distance to the natal (or foster) colony[26,27]. The colony of recruitment may thus be significantly influenced by the distribution of colonies of varying sizes and by the distance from the colony of origin, independently of the possible preference of the individual.

Our simulations of both non-experimental and experimental studies randomly produced high proportions of significant parent-offspring regressions of colony size choice and showed that these frequencies are highly sensitive to types of colony distributions (Figure 1a and 1c). More importantly, our simulations suggest that experimental data may be even more exposed to the two fallacies (Figure 1a).

Our simulations show that it is necessary to avoid both pitfalls to conclude that a behavioral trait such as habitat selection has a heritable genetic component in the population. Our final simulations suggest that the two pitfalls can be avoided by designing experiments that provide individuals with the same set of options. The standard method for estimating heritability is the partial cross-foster experiment[39,40], as was performed in the cliff swallow study. In this design, half of the nestlings were swapped between nests of small and large colonies creating two types of offspring dyads, full sibs in nests in different colonies, and foster sibs in the same nests.

Although the goal of this design is to use these dyads to perform two pair-wise tests in a single analysis, this test was not reported in the cliff swallow study. However, being raised in the same nest provides foster siblings with the same set of options, which according to our simulations would make their comparison immune to both the spatial fallacy and RTM. In contrast, full siblings raised in different colonies have different sets of options, making any comparison between them susceptible to both pitfalls. The effect of these fallacies should increase the actual differences in colony size choice between full siblings, thus biasing the full analysis and, consequently, estimates of heritability.

Our simulations suggest that only comparisons between individuals with the same set of options are immune to the two pitfalls. Providing the same options may be achieved by fostering all offspring within a single foster colony. This method clearly avoids the spatial fallacy. However, while our simulations show that this also avoids the RTM fallacy when working with ranks (Figure 4), it is nevertheless possible that it only avoids RTM when using the most frequently occurring colony size as the common foster colony. Unfortunately, actual colony sizes were not reported in the cliff swallow study, which did not allow us to simulate them.

However, the logic of RTM allows us to predict that using the most frequently occurring colony size as the common foster colony will lead to a non-significant proportion of spurious heritabilities. Further simulations may demonstrate whether experimenters can be more flexible by being able to select a colony of any size as the common foster colony. Pending such simulations we propose selecting a colony of the most frequently occurring size.

Methods

Our main goal was to determine whether parent-offspring regressions produced by cross-foster experiments can be generated by randomly simulating Colony size choice. Brown and Brown's[14] methods and results were reported in considerable detail, allowing us to use their published data to simulate parent-offspring regressions of colony size choice. Throughout this paper unless otherwise noted, references to the cliff swallow study are of[14]. As in that study, we used breeding colony size choice as the phenotypic trait.

Our individual-based simulations produced a null model of the expected statistical significance of parent-offspring regressions when accounting for the spatial distribution of colonies of various sizes. We used the colony size ranks data and the number of recruits provided in Figure 2 of the cliff swallow study to simulate parent-offspring regressions. We were unable to simulate distributions of real colony sizes because this was not described in the cliff swallow study. In that study, a total of 721 chicks were recovered after they had been cross-fostered in the previous year between colonies of various sizes. After simulating the 721 recruits in each run, we estimated heritabilities by separately regressing recruitment colony size against both birth and foster colony size. These simulations are based on the logic of[14] that positive regressions to the birth colony and negative regressions to the foster colony support the existence of a genetic component in colony size choice.

We ran 2,000 Monte Carlo simulations for each set of parameters. A new distribution of colonies was recreated for each run. We tabulated the p-values of the generated parent-offspring regressions using the standard significance threshold of 0.05 (white bars in Figures 1 and 4a). To further examine whether the extremely high significance of the cliff swallow regressions (in which 8 out of 9 were significant, including four with p-values under 0.0001 and one that was negative to the rearing colony) could be generated randomly, we also used the actual p-values reported in Table 1 in[14] as significance thresholds (black bars in Figures 1 and 4a). We adopted[14] prediction that regressions of recruitment colony size should be positive to the birth and negative to the foster colony size in tabulating only simulated regressions that met these criteria. However, our simulations sometimes also generated substantial proportions of significant regressions of the unpredicted sign (see for instance Figure 1c).

Spatial distribution of colonies. In the cliff swallow study, colonies were aggregated in five clusters ranging from 11 to 25 colony sites. The largest cluster in a single year comprised 21 active colonies. We therefore simulated 21 colonies with size ranks ranging from 1 (largest) to 21 (smallest). We used a linear lattice of 150×5 cells to represent the fact that cliff swallows breed along rivers, and a square lattice of 50×50 cells, which may better represent the habitat of most species, including barn swallows and lesser kestrels.

We then simulated four different types of distributions of colonies of varying size ranks (examples of the simulated distributions are provided in the Supplementary Online Material). In all four distributions, colony positions were drawn randomly from all cells of the lattice, with a given cell containing at most one colony. We then assigned each selected cell (*i.e.*, colony position) a size rank according to four distributions. The first three distributions represented the fact that in nature, large colonies tend to be far apart[35–38].

In the **BigFar5** distributions, the five largest colonies were randomly assigned to some of the selected colony positions so that they were at least 20 cells apart. The remaining colonies were then assigned so that the smallest colony was closest to the largest colony, the second smallest near the second largest, and so on until the fifth smallest colony. The process was then reiterated for the next five smallest colonies so that the 6th smallest colony was in the remaining location closest to the largest, and so on until all colonies were placed. This type of distribution was designed to generate the maximum contrast in the sizes of neighbouring colonies in order to produce the highest probabilities of spurious significant experimental regressions. The purpose of this exercise was to explore the range of probabilities of producing spurious regressions when not accounting for the two pitfalls. By placing smaller colonies next to the five largest ones and assuming the random walk process of recruitment (see next section), individuals were likely to recruit to colonies of substantially different sizes than the ones from which they fledged. This was expected to generate negative regressions between recruitment and fledging colonies. In simulating cross-fostering experiments, the fledging colony was the foster colony, while in non-experimental situations the fledging colony was the birth colony.

In the **BigFar5Random** distributions, ranks of the five largest colonies were also at least 20 cells apart, whereas the ranks of the remaining colonies were assigned randomly.

The **BigFarHalf** distributions corresponded to the BigFar5 distributions in which colonies of the largest half of the distribution were separated by at least seven cells. The remaining smaller colonies were then placed so that the smallest was closest to the largest, the second smallest closest to the second largest, etc.

Finally, we simulated **Random** distributions which were totally random with regard to colony position and size. We designed the first three colony distributions to explore the impact of different kinds of spatial structuring of colony sizes and the random distribution as a null model to provide a basis of comparison with a non-structured distribution. This null model is conservative because our algorithm inevitably generates some outputs with some structuring. In the absence of any bias, we expect 5% of the simulations to generate significant regressions.

Recruitment. We simulated the philopatric recruits of the cliff swallow study by recruiting them to their birth colony. In contrast, non-philopatric individuals recruited according to a "random walk" strategy in which recruits returned to their fledging colony and then moved randomly within the lattice. At each step, all non-philopatric individuals had an equal probability of moving to any of the eight adjacent cells until they entered a cell containing a colony, to which they recruited. Simulated non-philopatric birds were not allowed to recruit to their fledging colony. This algorithm imitates the diffusion process in physics wherein molecules move randomly, and allowed us to account for the spatial distributions of colonies without assuming any process of choice by the birds. It also imitates natural situations such as when newly fledged birds explore their environment starting from their fledging location before migrating.

Selecting experimental colonies. To compare how the results of cross-foster experiments may be influenced by the ways that researchers select experimental

colonies, we simulated an experimental design in which the birth and recruitment colonies were randomly chosen. We then simulated another experimental design in which we systematically selected birth and foster colonies in order to cover the full range of contrasts in colony size ranks. The highest contrast was between the largest and smallest colony size ranks of 1 and 21, and the lowest contrast was between colonies of nearly the same intermediate rank, *i.e.*, 10 and 12.

Providing common options. Our final goal was to explore whether fostering all nestlings to a common colony may avoid RTM and the spatial fallacy by providing them with the same set of choices when dispersing from the same location. This method resembles a common garden experiment in which all individuals are fostered into the same location in order to apportion genetic and environmental effects on the phenotype[41,42]. Our proposed "common options" experiment is designed to additionally provide all individuals with the identical set of opportunities to disperse to any location.

In a first of two sets of simulations, we placed all fostered young into the median sized colony. In the second set, we performed the same simulations separately for foster colonies of each size rank. A design avoiding the two pitfalls should lead to 5% of significant parent-offspring regressions, of which only half (2.5%) should be positive to the birth and negative to the foster colony sizes in one-tailed tests.

All our simulations are based on ranks in colony size. We did not simulate distributions of real colony sizes because this was not described in the cliff swallow study. We reckon that such distributions would be bell shaped with the median colony also being a colony of a frequently occurring size. In such conditions the effect of RTM should be increased, which could change the shape of Figure 4 b, in having higher proportions of significant regressions when using a less frequently occurring colony as the common foster colony.

1. Mousseau, T. A. & Roff, D. A. Natural selection and the heritability of fitness components. *Heredity.* **59**, 181–197 (1987).
2. Mousseau, T. A. & Fox, C. W. The adaptive significance of maternal effects. *Trends Ecol Evol.* **13**, 403–407 (1998).
3. Merilä, J. & Sheldon, B. C. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity.* **83**, 103–109 (1999).
4. Merilä, J. & Sheldon, B. C. in *Current Ornithology* 16 (eds. Val Nolan, J. & Thompson, C. F.) 179–255 (Kluver Academic/Plenum Press, New York, 2001).
5. Wilson, A. J., Gelin, U., Perron, M.-C. & Réale, D. Indirect genetic effects and the evolution of aggression in a vertebrate system. *Proc R Soc Lond B Biol Sci.* **276**, 533–541 (2009).
6. Clobert, J., Danchin, É., Dhondt, A. A. & Nichols, J. D. (eds.) *Dispersal* (Oxford University Press, New York, 2001).
7. Weiss, A. & King, J. E. The heritability of personality factors in zoo chimpanzees *Pan troglodytes*. *Behav Genet.* **28**, 484–485 (1998).
8. Dingemanse, N. J., Both, C., Drent, P. J., van Oers, K. & van Noordwijk, A. J. Repeatability and heritability of exploratory behaviour in great tits from the wild. *Anim Behav.* **64**, 929–938 (2002).
9. Dingemanse, N. J., Both, C., van Noordwijk, A. J., Rutten, A. L. & Drent, P. J. Natal dispersal and personalities in great tits (*Parus major*). *Proc R Soc Lond B Biol Sci.* **270**, 741–747 (2003).
10. Sinn, D. L., Apiolaza, L. A. & Moltschaniwskyj, N. A. Heritability and fitness-related consequences of squid personality traits. *J Evol Biol.* **19**, 1437–1447 (2006).
11. Réale, D., Martin, J., Coltman, D. W., Poissant, J. & Festa-Bianchet, M. Male personality, life-history strategies and reproductive success in a promiscuous mammal. *J Evol Biol.* **22**, 1599–1607 (2009).
12. van Oers, K., Drent, P. J., de Goede, P. & van Noordwijk, A. J. Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proc R Soc Lond B Biol Sci.* (2013).
13. Charmantier, A., Keyser, A. J. & Promislow, D. E. L. First evidence for heritable variation in cooperative breeding behaviour. *Proc R Soc Lond B Biol Sci.* **274**, 1757–1761 (2007).
14. Brown, C. R. & Bomberger Brown, M. Heritable basis for choice of group size in a colonial bird. *Proc Nat Acad Sci USA.* **97**, 14825–14830 (2000).
15. Brown, C. R., Stutchbury, B. J. & Walsh, P. D. Choice of Colony Size in Birds. *Trends Ecol Evol.* **5**, 398–403 (1990).
16. Brown, C. R. & Bomberger Brown, M. in *Current Ornithology* (eds. Val Nolan, J. & Thompson, C. F.) 1–81 (Kluvwer Academic/Plenum Publishers, New York, 2001).
17. Jovani, R., Mavor, R. & Oro, D. Hidden patterns of colony size variation in seabirds: a logarithmic point of view. *Oikos.* **117**, 1774–1781 (2008).
18. Brown, C. R. & Brown, M. B. *Coloniality in the Cliff Swallow. The Effect of Group Size on Social Behavior* (University of Chicago Press, Chicago, 1996).
19. Krause, J. & Ruxton, G. D. *Living in Groups* (Oxford University Press, Oxford, 2002).
20. Magrath, M. J. L. *et al.* Seasonal decline in reproductive performance varies with colony size in the fairy martin. *Petrochelidon ariel. Behav Ecol Sociobiol.* **63**, 661–672 (2009).
21. Danchin, É. & Wagner, R. H. The evolution of coloniality: the emergence of new perspectives. *Trends Ecol Evol.* **12**, 342–347 (1997).
22. Møller, A. P. Parent-offspring ressemblance in degree of sociality in a passerine bird. *Behav Ecol Sociobiol.* **51**, 276–281 (2002).
23. Serrano, D. & Tella, J. L. The role of despotism and heritability in determining settlement patterns in the colonial lesser kestrel. *Am Nat.* **169**, E53–E67 (2007).
24. Dugatkin, L. A. Cliff swallows and the power of behavioral ecology. *Proc Nat Acad Sci USA.* **98**, 16–17 (2001).
25. van Noordwijk, A. J. Problems in the analysis of dispersal and critique on its "heritability" in the great tit. *J Anim Ecol.* **53**, 533–544 (1984).
26. Serrano, D., Carrete, M. & Tella, J. L. Describing dispersal under habitat constraints: A randomization approach in lesser kestrels. *Basic Appl Ecol.* **9**, 771–778 (2008).
27. Coulson, J. C. & Nève de Mévergnies, G. Where do young Kittiwakes *Rissa tridactyla* breed, philopatry or dispersal? *Ardea.* **80**, 187–197 (1992).
28. Galton, F. Regression toward mediocrity in hereditary stature. *J Antrhopol Inst.* **15**, 246–263 (1886).
29. Stigler, S. M. Statistics on the Table: *The History of Statistical Concepts and Methods* (Harvard University Press Cambridge, Mass, 1999).
30. Barnett, A. G., van der Pols, J. C. & Dobson, A. J. Regression to the mean: what it is and how to deal with it. *In J Epidemiol.* **34**, 215–220 (2005).
31. Kelly, C. & Price, T. D. Correcting for regression to the mean in behavior and ecology. *Am Nat.* **166**, 700–707 (2005).
32. Greenwood, P. J., Harvey, P. H. & Perrins, C. M. The role of dispersal in the great tit (*Parus major*): the causes, consequences and heritability of natal dispersal. *J Anim Ecol.* **48**, 123–142 (1979).
33. Korner-Nievergelt, F. *et al.* Improving the analysis of movement data from marked individuals through explicit estimation of observer heterogeneity. *J Avian Biol.* **41**, 8–17 (2010).
34. van Noordwijk, A. J. Measuring dispersal as distance-dependent recruitment rates: testing the performance of DDRR on simulated data. *J Ornithol.* **152**, S239–S249 (2011).
35. Furness, R. W. & Birkhead, T. R. Seabird colony distributions suggest competition for food supplies during the breeding season. *Nature.* **311**, 655–656 (1984).
36. Birkhead, T. R. & Furness, R. W. in Behavioural Ecology. *Ecological Consequences of adaptive Behaviour* (eds. Sibly, R. M. & Smith, R. H.) 145–167 (Blackwell Scientific Publications, Oxford, London, 1985).
37. Furness, R. W. & Monaghan, P. *Seabird Ecology* (Blackie, Glasgow and London, 1987).
38. Lewis, S., Sherratt, T. N., Hamer, K. C. & Wanless, S. Evidence of intra-species competition for food in a pelagic seabird. *Nature.* **421**, 816–819 (2001).
39. Falconer, D. S. *Introduction to quantitative genetics* (Longman, New York, 1981).
40. Falconer, D. S. & Mackay, T. F. C. *Introduction to quantitative genetics* (ed. ed, t.) (Longman, Harlow, Essex UK, 1996).
41. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol Let.* **7**, 1225–1241 (2004).
42. Monty, A., Lebeau, J., Meerts, P. & Mahy, G. An explicit test for the contribution of environmental maternal effects to rapid clinal differentiation in an invasive plant. *J Evol Biol.* **22**, 917–926 (2009).

Acknowledgments

We thank Donald Blomqvist, Fred Bookstein, Simon Blanchet, Benoit Pujol, Robert E. Ricklefs and Arie van Noordwijk for commenting on various versions of this paper and Dieter Ebert far a stimulating discussion. This work was supported by the French Laboratory of Excellence project "TULIP" (*ANR -10-LABX-41; ANR-11-IDEX-0002-02*) to E.D. and by the Austrian Academy of Sciences to R.H.W.

Author contributions

E.D. and R.H.W. designed the study and E.W. created the code for the simulations. E.D. performed the simulations and tabulated the results. The paper was written by E.D. and R.H.W.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

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

How to cite this article: Danchin, E., Wajnberg, E. & Wagner, R.H. Avoiding pitfalls in estimating heritability with the common options approach. *Sci. Rep.* **4**, 3974; DOI:10.1038/srep03974 (2014).