

Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data

Studies of populations in the wild can provide unique insights into the forces driving evolutionary dynamics. This themed issue of *Proc. R. Soc. B* focuses on new developments in long-term analyses of animal populations where pedigree information has been collected. These address fundamental questions in evolutionary biology concerning the genetic basis of phenotypic diversity, patterns of natural and sexual selection, the occurrence of inbreeding and inbreeding depression, and speciation. Contributions include the analysis of evolutionary responses to climate change, exploration of the genetic basis of senescence, the exploitation of advances in molecular genetic technology, and reviews of developments in quantitative genetic methodology. We discuss here common themes, specific problems and pointers for future research.

Keywords: evolution; pedigree; selection; quantitative genetics

For evolutionary biologists interested in what goes on outside the laboratory or farmyard, it is fortunate that in several different places around the world, at several different times, researchers initiated individual based studies of wild animal populations and then kept these studies going for decade after decade. The data generated by such long-term projects have turned out to be a goldmine for a host of different fields of zoology. In this special issue, we focus on the insights they provide into evolutionary dynamics. The papers have a common theme of exploring the use of long-term data in the cases where relatedness between individuals is known and so pedigrees can be constructed. We highlight here the value of that pedigree data.

Evolutionary change requires a combination of two key ingredients: selection altering the distribution of phenotypes, and heritable genetic variance underlying the phenotypic distribution such that the changes due to selection are passed on to subsequent generations. Analyses of both of these aspects require pedigree data: firstly, the estimates of individual breeding success that pedigrees provide can be linked to measures of phenotypic traits to determine selection and, secondly, quantifying the phenotypic covariance of relatives gives an indication of the genetic basis of a trait (the field of quantitative genetics). Furthermore, by providing information on mating patterns, pedigree data can also provide valuable insights into two other prominent branches of evolutionary biology, namely studies of the impact and avoidance of inbreeding, and studies of speciation.

The last decade has witnessed a rapid increase in activity in this area, and the aim of this themed issue is to showcase the recent developments in the field. The great majority of the studies presented here would probably not have been possible 10 years ago, for various reasons. In some cases, this is simply because datasets would not have been large enough: for example, the key events such as hybridization (Svedin *et al.* 2008) or extreme mating patterns (Grant & Grant 2008) may be rare, requiring the accumulation of sufficiently large datasets before sensible analyses are feasible. Secondly, advances in molecular technology have revolutionized the availability of

molecular genetic data and hence the ease with which parentage (typically paternities) can be reliably assigned, and hence pedigrees constructed (Pemberton 2008). Thirdly, the study of quantitative genetics in the wild has been revitalized by the adoption of more complex statistical techniques that allow full exploitation of multi-generational, complex, unbalanced pedigrees, in particular the ‘animal model’. These techniques have a long history in animal breeding (Henderson 1950; Thompson 2008), becoming practical during the 1980s as sufficient computing power became available, and it is perhaps not to the credit of evolutionary biologists that it took until 1999 to make use of the substantially more efficient and powerful approach they offer (Kruuk 2004). More generally, the use of mixed models to incorporate the multiple strata present in biological data has now become a commonplace, and has facilitated much more comprehensive analyses of large-scale datasets.

Finally, there is increasing awareness of the extent to which ecological conditions drive many aspects of evolutionary dynamics. Rather than being a statistical nuisance requiring correction as best as possible, and ultimately therefore a severe disadvantage of field studies, environmental effects and their interactions with evolutionary processes are increasingly appreciated as fundamentally important and interesting in their own right. The ability to incorporate real-world variability can therefore become a strength rather than a weakness of the studies of natural populations. As an illustration of this awareness, the studies presented in this issue explore the effects of ecological or environmental heterogeneity on: genetic variance (Brommer *et al.* 2008), natural and sexual selection (Cockburn *et al.* 2008; Sinervo & McAdam 2008), life-history trade-offs (Gillespie *et al.* 2008), and inbreeding (Szulkin & Sheldon 2008). Growing awareness of the effects of anthropogenic climate change has also greatly encouraged interest in the impact of changing environmental conditions (Visser 2008). Thus, there has been a marked shift towards a realization that analyses exploring the interactions between environmental conditions and key evolutionary processes in natural populations are both important and feasible—in part due

One contribution of 18 to a Special Issue ‘Evolutionary dynamics of wild populations’.

to the development of appropriate statistical methodology with which to model such effects (Nussey *et al.* 2007). Parallel arguments exist for the exploration of effects of ageing, where the age of an individual is treated as the environment in which a trait is expressed. Using this approach, evolutionary theories of the genetic basis of senescence are now being tested in wild populations, for example, through analyses for interactions with inbreeding (Keller *et al.* 2008) and genetically based trade-offs between performance at different stages of life (Nussey *et al.* 2008).

However, working with uncontrolled, unmanipulated populations experiencing natural rather than artificial selection also has its drawbacks. Several issues raised their heads repeatedly in the various contributions presented here, and no doubt will be familiar to anyone working in the field.

The first limitation to the analyses of natural selection is, because the essence of what we are interested in is the representation of genes in future generations, much depends on the concept of individual fitness. Many studies consider only a single component of fitness, no doubt a valuable approach, but it is never clear which other aspects of the complete picture may be missed. Estimating 'total' individual fitness is tricky, as is evidenced by the variety of measures employed in the literature, ranging from the very simple (e.g. lifetime production of offspring) to the more complex (e.g. instantaneous contribution to population growth; Coulson *et al.* 2006). Different measures can give different results and will also have different statistical properties; the diversity reflects what is still a lack of consensus on the best way to quantify fitness.

A second limitation is the breakdown of theoretical expectations for the direction and magnitude of evolutionary response (i.e. the uni- or multivariate breeder's equation). Artificial selection of known magnitude reliably produces predictable responses (Falconer & Mackay 1996), but observed natural selection rarely does (Merilä *et al.* 2001). While it is possible to envisage a range of explanations as to why this might be so (e.g. see review in Merilä *et al.* 2001), one inescapable aspect is the action of correlated but unmeasured selection. This may be via selection acting on other correlated traits, which are effectively invisible if not incorporated into analyses, or simply because viability selection on the focal trait prior to census alters that trait's distribution; the dead are similarly invisible. Hadfield (2008) quantifies the dramatic effect that such invisibility can have on estimation procedures, and the concept is explored empirically with long-term data from a lizard population (Sinervo & McAdam 2008).

There are several other statistical concerns which arise frequently with data from natural populations. While sample sizes and lengths of studies might be remarkable in terms of the number of hours of fieldwork required to accumulate them, they are small (usually in the hundreds or low thousands) relative to those typical of pedigreed livestock populations (which may reach the millions) and some laboratory experiments. Nor, unlike the latter, can they usually be designed to get most information from the data, using known selection criteria and equal family sizes. Add to this interactions with changing environmental conditions, and a scarcity of data can become a serious constraint on the ambitions of particular models. Concerns about lack of statistical power are therefore inevitable. In many cases, careful analyses can do no more

than to conclude that while there is some support for the phenomenon being tested, the current data have insufficient power to distinguish between relevant hypotheses (e.g. Brommer *et al.* 2008; Keller *et al.* 2008; Nussey *et al.* 2008; Sinervo & McAdam 2008). Many such studies aim to quantify the magnitude of some form of variance component, for example additive genetic variance or differences between individuals in reaction norms in a changing environment. In the vast majority of such cases, it seems highly unlikely that a null hypothesis of absolutely zero variance is actually true. This underlines the need for careful choice of wording when reporting null results: lack of a significant result may just be lack of power. Furthermore, with some data structures, the lack of statistical significance of particular variance components is not a sufficient reason to drop them from a model. In many cases, additive genetic variance (V_A) will be badly overestimated from poorly specified models (Kruuk & Hadfield 2007): for example, Ovaskainen *et al.* (2008) illustrate the impact that dominance variance may have on estimates of V_A . (The ability to test for dominance variance using natural pedigrees remains unexplored, but experience with populations of livestock species does not make us optimistic that a resolution is probable.) Such problems are exacerbated in multitrait situations. All these issues highlight the need for careful model specification, and also the recognition that while adequate models may be found, the 'right' model may be unknowable.

Despite the above points, consideration of future avenues of research suggests that there is still much to be gained from mining the data accumulated from long-term pedigrees. Exploration of several important current challenges in evolutionary biology are still in their infancy, including the analysis of sexually antagonistic genetic variance (Poissant *et al.* 2008), the genetic basis of responses to changing environmental conditions (Brommer *et al.* 2008; Visser 2008), exploration of the genetic basis of senescence (Keller *et al.* 2008; Nussey *et al.* 2008) and the exploitation of rapid advancements in molecular genetic technology (Slate 2008). The questions being asked of datasets from natural populations have clearly moved considerably beyond simply estimating the heritability of a trait or the selection pressures to which it is subject. As more results accumulate, we will be able to look for generalizations: for example, in how changing environmental conditions alter both the expression of genetic variance (Charmantier & Garant 2005) and patterns of natural selection. Nevertheless, if prevailing environmental and ecological conditions are as important as suggested by current results, generalizations across different populations in different environments may become increasingly difficult.

We are still relatively restricted in the taxa being considered, with birds and mammals (especially passerines and ungulates) continuing to dominate (this issue and see reviews in Kruuk 2004; Nussey *et al.* 2007). Until marker-based estimates of quantitative genetic parameters or entire marker-based recovery of a pedigree become more feasible, many taxa such as fish or invertebrates for which individual monitoring is relatively more difficult (though not impossible; e.g. Bonduriansky & Brassil 2005) are likely to be under-represented. We are also fairly conservative in the traits that are analysed, returning repeatedly to morphological variables such as body size or

secondary sexual characteristics and life-history variables such as fecundity or timing of breeding. Whole new avenues remain to be opened up through the greater exploration of alternative traits, such as behavioural or physiological variables.

However, with the rapidly increasing feasibility and reducing costs of genotyping, it becomes increasingly possible to accurately determine the relationships and construct pedigrees including quite distantly related individuals. The current consensus seems to be that the value of molecular data lies in improved determination of relatedness between individuals in a population (i.e. more accurate construction of a pedigree) rather than the possibility of bypassing pedigree construction and estimating quantitative genetic parameters directly from comparisons of phenotypic and marker data (Frentiu *et al.* 2008; Pemberton 2008), but this may change. The availability of large numbers of markers will also facilitate the mapping of loci of adaptive significance (Slate 2008). The main constraints may then lie in obtaining the funding for sufficient numbers of genotypes. Furthermore, while such information will facilitate some studies, in many cases it is likely to be available only on collateral relatives, and thus is unlikely to replace long-term studies which can be used to determine evolutionary constraints or change.

Statistically, there is also clearly much to explore. Treatment of multivariate rather than univariate phenotypes has the potential to radically alter interpretations (Blows 2007), and full characterization of genetic covariances and correlations across different combinations of traits is essential for tests of several key hypotheses (e.g. Nussey *et al.* 2008; Poissant *et al.* 2008). There is also increasing appreciation that the scenarios to which evolutionary biologists apply the animal modelling techniques may be profoundly different from the more controlled arrangements for which they were designed. To this end, Bayesian approaches may provide superior means of incorporating issues which have rather been swept under the carpet to date, such as assessments of uncertainty, more complex statistical distributions or impacts of selection (Hadfield 2008; Ovaskainen *et al.* 2008).

We would like to conclude by thanking the numerous people involved with the production of this special issue. Firstly, we thank all the authors for their contributions (including those disappointed by the outcome) and for their patience during the rounds of revisions. All papers were independently refereed and both the referees and the editors made tough demands in places. Secondly, we are as ever extremely grateful for the time and effort of the referees. Thirdly, many thanks to the tireless production team of *Proc. R. Soc. B*. Finally, as a research community, we should be collectively grateful for the essential work done by those who initiated and maintained the studies involved—they may not have known at the time how long they would run for, or what the data might be used for, but without their foresight most of what is presented here would not be possible. The longest ongoing study in this themed issue has been running since 1947 (Szulkin & Sheldon 2008) and in several cases, studies have been maintained only by a continuous flow of short-term funding. Uninterrupted data collection is obviously essential for data quality, but guaranteeing it is a process that becomes increasingly uncertain and time consuming

as funding budgets get tighter. We hope that raised awareness of the long-term value and increasing returns from such datasets may help in maintaining long-term studies into the future.

L. E. B. Kruuk*

W. G. Hill

March 2008

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK

*Author for correspondence (loeske.kruuk@ed.ac.uk).

REFERENCES

- Blows, M. W. 2007 A tale of two matrices: multivariate approaches in evolutionary biology. *J. Evol. Biol.* **20**, 1–8. (doi:10.1111/j.1420-9101.2006.01164.x)
- Bonduriansky, R. & Brassil, C. E. 2005 Reproductive ageing and sexual selection on male body size in a wild population of antler flies (*Protophihila litigata*). *J. Evol. Biol.* **18**, 1332–1340. (doi:10.1111/j.1420-9101.2005.00957.x)
- Brommer, J. E., Rattiste, K. & Wilson, A. J. 2008 Exploring plasticity in the wild: laying date–temperature reaction norms in the common gull *Larus canus*. *Proc. R. Soc. B* **275**, 687–693. (doi:10.1098/rspb.2007.0951)
- Charmantier, A. & Garant, D. 2005 Environmental quality and evolutionary potential: lessons from wild populations. *Proc. R. Soc. B* **272**, 1415–1425. (doi:10.1098/rspb.2005.3117)
- Cockburn, A., Osmond, H. L. & Double, M. C. 2008 Swingin' in the rain: condition-dependence and sexual selection in a capricious world. *Proc. R. Soc. B* **275**, 605–612. (doi:10.1098/rspb.2007.0916)
- Coulson, T., Benton, T. G., Lundberg, P., Dall, S. R. X., Kendall, B. E. & Gaillard, J.-M. 2006 Estimating individual contributions to population growth: evolutionary fitness in ecological time. *Proc. R. Soc. B* **273**, 547–555. (doi:10.1098/rspb.2005.33357)
- Falconer, D. S. & Mackay, T. F. C. 1996 *Introduction to quantitative genetics*. Essex, UK: Longman.
- Frentiu, F. D., Clegg, S. M., Chittock, J., Burke, T., Blows, M. W. & Owens, I. P. F. 2008 Pedigree-free animal-models: the relatedness matrix reloaded. *Proc. R. Soc. B* **275**, 639–647. (doi:10.1098/rspb.2007.1032)
- Gillespie, D. O. S., Russell, A. F. & Lummaa, V. 2008 When fecundity does not equal fitness: evidence of an offspring quantity vs. quality trade-off in pre-industrial humans. *Proc. R. Soc. B* **275**, 713–722. (doi:10.1098/rspb.2007.1000)
- Grant, P. R. & Grant, B. R. 2008 Pedigrees, assortative mating and speciation in Darwin's finches. *Proc. R. Soc. B* **275**, 661–668. (doi:10.1098/rspb.2007.0898)
- Hadfield, J. D. 2008 Estimating evolutionary parameters when viability selection is operating. *Proc. R. Soc. B* **275**, 723–734. (doi:10.1098/rspb.2007.1013)
- Henderson, C. R. 1950 Estimation of genetic parameters. *Ann. Math. Stat.* **21**, 309–310.
- Keller, L. F., Reid, J. M. & Arcese, P. 2008 Testing evolutionary models of senescence in a natural population: age and inbreeding effects on fitness components in song sparrows. *Proc. R. Soc. B* **275**, 597–604. (doi:10.1098/rspb.2007.0961)
- Kruuk, L. E. B. 2004 Estimating genetic parameters in wild populations using the 'animal model'. *Phil. Trans. R. Soc. B* **359**, 873–890. (doi:10.1098/rstb.2003.1437)

- Kruuk, L. E. B. & Hadfield, J. D. 2007 How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* **20**, 1890–1903. (doi:10.1111/j.1420-9101.2007.01377.x)
- Merilä, J., Sheldon, B. C. & Kruuk, L. E. B. 2001 Explaining stasis: microevolutionary studies of natural populations. *Genetica* **112**, 119–222. (doi:10.1023/A:1013391806317)
- Nussey, D. H., Wilson, A. J. & Brommer, J. E. 2007 The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Biol.* **20**, 831–844. (doi:10.1111/j.1420-9101.2007.01300.x)
- Nussey, D. H., Wilson, A. J., Morris, A., Pemberton, J., Clutton-Brock, T. & Kruuk, L. E. B. 2008 Testing for genetic trade-offs between early- and late-life reproduction in a wild red deer population. *Proc. R. Soc. B* **275**, 745–750. (doi:10.1098/rspb.2007.0986)
- Ovaskainen, O., Cano, J. M. & Merilä, J. 2008 A Bayesian framework for comparative quantitative genetics. *Proc. R. Soc. B* **275**, 669–678. (doi:10.1098/rspb.2007.0949)
- Pemberton, J. M. 2008 Wild pedigrees: the way forward. *Proc. R. Soc. B* **275**, 613–621. (doi:10.1098/rspb.2007.1531)
- Poissant, J., Wilson, A. J., Festa-Bianchet, M., Hogg, J. T. & Coltman, D. W. 2008 Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proc. R. Soc. B* **275**, 623–628. (doi:10.1098/rspb.2007.1361)
- Sinervo, B. & McAdam, A. G. 2008 Maturational costs of reproduction due to clutch size and ontogenetic conflict as revealed in the invisible fraction. *Proc. R. Soc. B* **275**, 629–638. (doi:10.1098/rspb.2007.1084)
- Slate, J. 2008 Robustness of linkage maps in natural populations: a simulation study. *Proc. R. Soc. B* **275**, 695–702. (doi:10.1098/rspb.2007.0948)
- Svedin, N., Wiley, C., Veen, T., Gustafsson, L. & Qvarnström, A. 2008 Natural and sexual selection against hybrid flycatchers. *Proc. R. Soc. B* **275**, 735–744. (doi:10.1098/rspb.2007.0967)
- Szulkin, M. & Sheldon, B. C. 2008 Dispersal as a means of inbreeding avoidance in a wild bird population. *Proc. R. Soc. B* **275**, 703–711. (doi:10.1098/rspb.2007.0989)
- Thompson, R. 2008 Estimation of quantitative genetic parameters. *Proc. R. Soc. B* **275**, 679–686. (doi:10.1098/rspb.2007.1417)
- Visser, M. E. 2008 Keeping up with a warming world: assessing the rate of adaptation to climate change. *Proc. R. Soc. B* **275**, 649–659. (doi:10.1098/rspb.2007.0997)