A peer-reviewed version of this preprint was published in PeerJ on 27 October 2017.

View the peer-reviewed version (peerj.com/articles/3938), which is the preferred citable publication unless you specifically need to cite this preprint.

Inferring microevolution from museum collections and resampling: lessons learned from Cepaea

Malgorzata Ozgo 1, Thor-Seng Liew 2,3,4, Nicole B Webster 3,4, Menno Schilthuizen Corresp. 2,3,4

1 Dept. of Evolutionary Biology, Kazimierz Wielki University, Bydgoszcz, Poland
2 Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia
3 Institute Biology Leiden, Leiden University, Leiden, The Netherlands
4 Endless Forms Group, Naturalis Biodiversity Center, Leiden, The Netherlands

Corresponding Author: Menno Schilthuizen
Email address: menno.schilthuizen@naturalis.nl

Studies documenting Human-Induced Rapid Evolutionary Change (HIREC) routinely compare contemporary allele or morph frequency distributions with historical baselines. All too often, this involves the re-sampling of a population that was sampled at a single time point in the past. However, year-to-year fluctuations in magnitude and direction of evolutionary response may make such studies prone to erroneous conclusions, where long-term evolutionary trends are inferred from what in fact are short-term fluctuations. Here, we explore this problem by re-sampling three Dutch populations of the land snail Cepaea nemoralis, whose shell colour polymorphism is known to be under thermal and predatory selection. Each of these three populations was originally sampled in at least two different years in the past. We show that conclusions on evolutionary change are strongly dependent on which of the historical sample dates is used for comparison with the contemporary sample. Our study highlights the fact that year-to-year variation in allele frequencies may often be so strong that a simple two-point comparison is unreliable to detect long-term evolutionary trends.
Inferring microevolution from museum collections and resampling: lessons learned from *Cepaea*

Małgorzata Ożgo,¹ Thor-Seng Liew,²,³,⁴ Nicole B. Webster³,⁴ & Menno Schilthuizen²,³,⁴*

¹ Dept. of Evolutionary Biology, Kazimierz Wielki University, Al. Ossolinskich 12, 85-064, Bydgoszcz, Poland
² Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, the Netherlands
³ Institute Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden
⁴ Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

* menno.schilthuizen@naturalis.nl

Abstract

Studies documenting Human-Induced Rapid Evolutionary Change (HIREC) routinely compare contemporary allele or morph frequency distributions with historical baselines. All too often, this involves the re-sampling of a population that was sampled at a single time point in the past. However, year-to-year fluctuations in magnitude and direction of evolutionary response may make such studies prone to erroneous conclusions, where long-term evolutionary trends are inferred from what in fact are short-term fluctuations. Here, we explore this problem by re-sampling three Dutch populations of the land snail *Cepaea nemoralis*, whose shell colour polymorphism is known to be under thermal and predatory selection. Each of these three populations was originally sampled in at least two different years in the past. We show that conclusions on evolutionary change are strongly dependent on which of the historical sample dates is used for comparison with the contemporary sample. Our study highlights the fact that year-to-year variation in allele frequencies may often be so strong that a simple two-point comparison is unreliable to detect long-term evolutionary trends.
As humans rapidly alter environmental conditions, opportunities for natural populations to evolutionarily adapt to such changes are increasing (Schilthuizen & Kellermann, 2013). Several examples of such rapid, contemporary evolution are already available (e.g., Cheptou et al., 2008; Buswell et al., 2011). In many cases, evidence comes from the re-sampling of populations that have been sampled before, with historical data coming either from the literature or from natural history collections (Suarez & Tsutsui, 2004; Schilthuizen et al., 2015). One frequently under-appreciated factor in such analyses is the temporal scale, which is important for three reasons. First, the time scale at which populations respond adaptively, depends on the organism’s generation time and generation overlap. Second, the time scale at which environmental changes take place may differ: gradual long-term trends may mask many short-term, more erratic changes (Siepielski et al., 2009). Third, sampling and re-sampling itself has a temporal scale that may or may not match the scale at which environmental changes and adaptive responses take place: it may consist of just a few time points, several, or even hundreds, of generations apart, or it may consist of an uninterrupted time series.

One system frequently employed for studies of contemporary adaptive evolution is shell colour evolution (often driven by thermal selection) in the land snails *Cepaea nemoralis* and *C. hortensis*. Shell colour in these species derives from well-studied genetic polymorphisms; the shells can be yellow, pink, or brown, and bear up to five dark brown spiral bands; effectively unbanded (eU) are shells with at least the two upper bands missing. Selection by climate and visually hunting predators can result in shifts in the frequencies of colour and banding morphs (reviews in, e.g., Cook 1998, Ożgo 2008). The animals are relatively sessile and easily sampled, and the species have been the subject of more than a century of investigations into local adaptation. For these reasons, several studies in *Cepaea* have made use of re-sampling to investigate temporal changes in shell colour frequencies in response to environmental change. These range from studies encompassing several years (e.g. Goodhart 1973, Arthur, Philips and
Mitchell 1993) to those spanning decades (e.g. Cook, Cowie and Jones 1999, Cameron 2001, Cameron and Pokryszko 2008, Cameron, Cook and Greenwood 2013). Most are geographically localized, but some encompass large areas. One of the largest citizen-science projects covered the whole of Europe (Silvertown et al. 2011).

In this paper, we further illustrate the problems that may ensue when assessing the degree of evolutionary change by insufficient re-sampling of a single population. We identified several large samples of *Cepaea nemoralis* suitable for this purpose in the collection of Naturalis Biodiversity Center in Leiden, the Netherlands. Then, we resampled these populations in 2010. We examined the frequency changes in three aspects of shell morph variation, namely, (1) shell ground colour; (2) shell spiral banding; and (3) relative frequency of yellow, effectively unbanded shells. (Details of scoring schemes are explained in the Materials and Methods). We show that the conclusions depend very strongly on the year with which our recent samples are compared. This suggests that, as shown in time series in a related species, *Theba pisana* (Johnson 2011), year-to-year evolutionary changes are often so large that they may mask, and lead to misleading interpretations of, long-term evolutionary trends.

**Materials and Methods**

**Historical Baseline Data**

We used the dry Mollusca collection of Naturalis Biodiversity Center to locate large samples of *Cepaea nemoralis* from the Netherlands with locality information precise enough to make re-sampling possible. We identified three such sets of samples (at the time of writing, these samples had not yet received collection numbers; however, they were held in the special “large Cepaea sample” section):

(1) Lobith. Shells collected by H. Wolda and students on 20th April 1960 (N=368), 27-28th April 1961 (N=398), and 10-14th April 1962 (N=1657) at 13 positions along a 400-m section of dike on the river bank along the Rhine (51.859° N 6.085° E). The samples were used for a study on stability of a steep c. 20-m-wide cline located along the
section, where unbanded yellow shells increase in frequency at the expense of banded yellow and unbanded “red” shells (Wolda, 1969). Wolda (1969) reports on a total of 26,230 individuals, and writes that most samples consisted of mostly living snails and a small proportion of dead shells. Since Wolda states that all living snails were returned to their collection localities after scoring, we conclude that the Naturalis samples (N=2423) consist only of snails that were collected as empty shells. Wolda (1969) also reports that the vegetation along the dike grades from a dense river-dune vegetation on sand, rich in nutrition and calcium, to more open river-dune vegetation on coarse, rich, and dry sand with calcium.

(2) Empe. Shells collected by J.C. van Heurn in May 1915 (N=468) and July 1951 (N=340) at Empe estate (52.145° N 6.142° E). The sample of 1915 is labelled as follows: “Collected at random in one site on the main road from Zutphen to Voorst, in the Zutphen municipality, directly opposite the manor “Empe”. The collection consists of 451 adults and 17 juveniles. [...] The habitats are the elm stems as well as the roadside-berges. In the verges there were no nettles or generally tall herbs, but there was a row of elm-bushes, mixed with some hawthorn, grasses and other low herbs [translated from Dutch by M.S.]”. The sample of 1951 is labelled, “Road south of Voorst. Opposite “Empe””. We conclude that both samples were taken at the exact same locality.

(3) Allemansgeest. Shells collected by W.C. van Heurn in summer 1942 (N=500) and 1943 (N=772) on the 150 x 50 m peninsula “Allemansgeest” at the confluence of Vliet and Korte Vliet, near Voorschoten (52.143° N 4.468° E). The location and samples were reported on by Van Heurn (1943, 1945). He describes the habitat as “a narrow spit of land, planted with willow, ash, and some alder, among which a wild herb vegetation appears in summer, consisting mostly of nettles [translated from Dutch by M.S.]”. He also reports on a set of samples from 1941, which we could not locate. The collected numbers reported for 1942 and 1943 were, respectively, 535 and 787, suggesting that a small number of individuals had been lost from the samples. We assume these to have been random subsets.

(A fourth set, from the vicinity of Eenrum, was already reported on previously; Ożgo & Schilthuizen, 2012.)
Resampling in 2010

We visited these exact locations in 2010. All sites were identifiable and still contained *Cepaea nemoralis*. At Lobith, the vegetation remained as described by Wolda (1969). Since snail densities were apparently much lower than in Wolda’s time, we did not sample Wolda’s individual sections, but instead pooled all material from the entire 400-m section of dike (20-22 June 2010). At Empe, no patch of vegetation exactly matched the description by Van Heurn, so we sampled from two different patches facing the façade of the manor house: Empe-1 (52.1457° N; 6.1435° E; reeds, nettles; 20-22 June 2010) and Empe-2 (52.1456° N; 6.1426° E; maple forest; 20-22 June 2010). At Allemansgeest, finally, the vegetation had become more park-like, with lawns, tall poplars, and only a few patches where a herbaceous layer was present. *Cepaea* densities were low, so we sampled multiple times, on 21 June, 22 August, and 13 September 2010. At all sites, we collected juveniles and adults, (fresh) dead and alive. Juveniles too small to judge the colour morph accurately were not included.

Shell Morph Scoring Scheme

Each individual from the museum samples and the recent samples was scored following the same scoring scheme as described in Cain and Sheppard (1954). We investigated three types of changes in shell morph frequencies between the historical baseline data and the resampling data: (1) changes in shell ground colour: yellow (Y), pink (P), and brown (B); (2) changes in banding categories: mid-banded (00300), three-banded (00345), five-banded (12345), and other banding; and (3) changes in yellow, effectively unbanded (YeU).

Data analysis

At each site, and for each pair of samples, we performed separate chi-square goodness-of-fit tests for each of the three types of morph change (banding, colour, and proportion of YeU) to examine whether there is evidence of a change in frequency between years (Supplementary Data 5). Although we recorded the frequency of all morph categories, we excluded categories from the goodness-of-fit tests when the morph frequency in the baseline year was equal to zero. The significance level of the
chi-square test was set at $p = 0.05$, and we tested for the following changes in frequency between the baseline and resampling years. Colour: yellow, pink, and brown; shell banding: mid-banded, three-banded and five-banded; YeU: Yellow-effectively unbanded. All the analyses were done in R v. 3.2.2, and the R script is provided in Supplementary Data 1.

Results

Full results are given in Figure 1 and Supplementary Data 2-4. We found no significant differences in colour ($\chi^2 = 2.38$, d.f. = 2, $p = 0.30$) or banding morph frequencies ($\chi^2 = 2.03$, d.f. = 2, $p = 0.36$) between the two patches sampled at Empe in 2010, so we pooled the data. Altogether, our 2010 sample sizes for Lobith, Empe, and Allemansgeest were $N = 131$, $N = 189$, and $N = 77$, respectively. Parts of these samples were stored as vouchers in the alcohol collection of Naturalis Biodiversity Center, under collection numbers RMNH.5004222 (Allemansgeest), RMNH.5004223-5004224 (Lobith) and RMNH.5004225-5004227 (Empe).

In Lobith 2010, there was a very high frequency of usually rare “other banding” snails, which constituted 26% of all shells and consisted mostly (12 of 13) of the 00045 form. Among the 1960, 1961, and 1962 samples there were statistically significant differences in colour, banding morph frequencies, and frequencies in yellow, effectively unbanded between consecutive years (Figure 1). For example, yellow effectively unbanded showed a drop in frequency from 54% in 1960 to 48% in 1961, followed by a significant increase to 61% in 1962. Consequently, the result of comparison with 2010 depends on which of the historic samples is considered. 1960 and 1961 give a highly significant difference in the YeU frequency, while 1962 does not. Similarly, 2010 does not show significantly different colour morphs compared with 1960 or 1961, but it does in comparison with 1962, mainly due to the proportion of brown shells.
In Empe, apparently no changes in any morph category occurred between 1915 and 2010. Strikingly, however, several colour morph frequencies in the 1951 sample differ significantly from both 1915 and 2010 (Figure 1). For example, the frequency of yellow decreased from 75% in 1915 to 66% in 1951, and increased again to 77% in 2010.

In Allemansgeest (Figure 1) there are no differences in the frequency of colour, banding, or yellow effectively unbanded morphs between the 1943 and 2010 samples. Between 1942 and 2010, however, significant differences in the frequency of banding and yellow effectively unbanded are observed. This reflects the change in morph composition that occurred between 1942 and 1943: differences in the frequency of banding and yellow effectively unbanded between these samples are significant and reflect an increase in mid-banded and a decrease in five-banded.

Discussion

Increasing evidence shows that adaptive response of populations to short term environmental changes can exceed changes observed over longer time periods (Siepielski et al. 2009, Bell 2010, Johnson 2011), and the results of the present study are most readily interpreted in this view. In Lobith, the proportion of yellow effectively unbanded (YeU), the character that responds strongly to visual (Cain and Sheppard 1954) and thermal selection (Ożgo 2011, Ożgo & Schilthuizen 2012), was significantly lower in 1961 than in 1960, and significantly higher in 1962 than in 1961. However, there was statistically significant difference between 1960 and 1962. Consequently, when the 2010 sample is compared with 1962, no changes are observed; when the same data are compared with 1961, a significant increase in YeU is recorded and suggests an adaptive genetic shift. Similarly, a significant increase in YeU was observed in Allemansgeest between 1942 and 2010, also suggesting a strong adaptive response. However, no such change was observed between 1943 and 2010; a significant increase in YeU occurred between 1942 and 1943. Possibly the most striking is the case of Empe, where the re-survey encompassed the longest time interval. A
significant increase in YeU was observed between 1915 and 1951, but the frequency of
YeU in 2010 was not significantly different from either 1915 or 1951. This suggests
cautions in comparing population states when sampling is carried out at long time
intervals.

The results of this study might be very different if we had only one historic sample
available from each locality. If we were comparing the results of the 2010 survey with
the surveys of 1915 in Empe, 1943 in Allemansgeest, and 1962 in Lobith, the
conclusion would be that no or very few observable changes occurred over those time-
intervals. If however, by coincidence, we had available the data collected in Empe in
1951, in Allemansgeest in 1942, and in Lobith in 1961, the conclusion might be that the
observed changes are significant and consistent, and indicate a response of
populations to contemporary climatic shift.

Obviously, when analyzing evolutionary changes, long time series of annual censuses
are the most reliable and informative. In polymorphic land snails such a study was
carried out by Cain et al. (1990), and subsequently re-analyzed by Bell (2010), and by
Johnson (2011). This approach, however, is limited to previously initiated and
maintained studies. These are regrettably few and generally they are incomparable as
the time periods, geographic locations, and taxonomic groups involved are too far apart.

Another reliable approach is to compare subdivided populations when time of
divergence is known (synchronic design sensu Hendry and Kinnison 1999). In Cepaea
nemoralis, this approach allowed us to infer rapid adaptive changes in response to
of divergence, however, is rarely known. Possibly the most promising approach is to
use a group of populations all sampled and resampled at the same time (e.g. Ożgo and
Schilthuizen 2012, Cameron et al. 2013). Here, the underlying assumption is that
organisms in different populations respond similarly to the same larger-scale selection
pressures; consistent changes over a number of populations are likely to reflect
directional processes.
This study shows that conclusions drawn from comparisons of single populations resurveyed after long time intervals can be misleading. Consequently, other studies where populations were surveyed only a couple of times can possibly lead to results that are difficult to interpret (e.g. Silvertown et al. 2011). In sum, our data might contribute to the increasing recognition of large changes in the states of populations over very short time scales. Such changes may result from natural selection in response to biotic or abiotic factors (e.g. changing weather conditions and/or fluctuating selection by predators), and from random events. We also draw attention to some methodological concerns of selection studies, and point to the value of long time series in ecological and evolutionary research.

Author Contributions
MO conceived the study, did the field work, carried out the shell scoring for field and museum samples, did the initial analyses, and wrote drafts for sections of the paper. MS conceived the study, did the field work, carried out the shell scoring for field and museum samples, and wrote the final draft of the paper. NW did the field work, revised drafts of the paper, and prepared the figure. LTS did the field work, refined the data analyses, and revised the draft of the paper.

Acknowledgements
We thank Jeroen Goud and Bram van der Bijl, collection managers of the Naturalis Mollusca collection for providing access to the historical specimens and for arranging the administration of the 2010 vouchers. Henrik Wolda, Rinny Kooi, Wilke van Delden, and Lucy Oosterhoff helped in obtaining additional information on historical collection localities.

References


Figure 1. Proportional variation in morph frequencies for all sites in all years. A, Shell ground colour morphs. B, Banding categories — Mid: mid-banded (00300); Three: three-banded (00345); Five: five-banded (12345); Other: other banding categories. C, Proportion of Yellow effectively unbanded (YeU). Numbers indicate total count. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. ¹Brown was excluded from all $\chi^2$ tests except Empe from 1915, and Lobith 1962-2010. ²Mid-banded was excluded from all Empe 1915 and Lobith 1960 $\chi^2$ tests. ³Mid-banded was excluded from Empe 1951-2010 $\chi^2$ test.
Supplementary Data 1. R scripts for the analyses.

Supplementary Data 2. Full details of the original and the resampled collections from Lobith, including a photo of the contemporary habitat.

Supplementary Data 3. Full details of the original and the resampled collections from Empe, including a photo of the contemporary habitats.

Supplementary Data 4. Full details of the original and the resampled collections from Allemansgeest, including a photo of the contemporary habitat.
Proportional variation in morph frequencies for all sites in all years.

Figure 1. Proportional variation in morph frequencies for all sites in all years. A, Shell ground colour morphs. B, Banding categories — Mid: mid-banded (00300); Three: three-banded (00345); Five: five-banded (12345); Other: other banding categories. C, Proportion of Yellow effectively unbanded (YeU). Numbers indicate total count. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. ¹Brown was excluded from all $\chi^2$ tests except Empe from 1915, and Lobith 1962-2010. ²Mid-banded was excluded from all Empe 1915 and Lobith 1960 $\chi^2$ tests. ³Mid-banded was excluded from Empe 1951-2010 $\chi^2$ test.