Theory Biosci. (2002) 121: 297–320

http://www.urbanfischer.de/journals/theorybiosc

From Haeckel to event-pairing: the evolution of developmental sequences

Olaf R. P. Bininda-Emonds[1,2] Jonathan E. Jeffery[1], Michael I. Coates[3], and Michael K. Richardson[1]

[1]Institute of Evolutionary and Ecological Sciences, Kaiserstraat 63, P. O. Box 9516, Leiden University, 2300 RA Leiden, The Netherlands
[2]Current address: Lehrstuhl für Tierzucht, Technical University of Munich, Alte Akademie 12, 85354 Freising-Weihenstephan, Germany
[3]Department of Organismal Biology & Anatomy, University of Chicago, 1027 East 57th Street, Chicago, Illinois 60637-1508, USA

Address for correspondence: Olaf R. P. Bininda-Emonds, Lehrstuhl für Tierzucht, Technical University of Munich, Alte Akademie 12, D-85354 Freising-Weihenstephan, Germany, e-mail: olaf.bininda@tierzucht.tum.de, phone: +49 (0) 81 61 71 37 41; fax: +49 (0) 81 61 71 31 07

Received: May 5, 2002; accepted: June 21, 2002

Key words: Developmental sequence, Haeckel, event-pairing, Normal Table, phylogeny

Summary: Development involves a series of developmental events, separated by transformations, that follow a particular order or developmental sequence. The sequence may in turn be arbitrarily subdivided into contiguous segments (developmental stages). We discuss the properties of developmental sequences. We also examine the differing analytical approaches that have been used to analyse developmental sequences in an evolutionary context. Ernst Haeckel was a pioneer in this field. His approach was evolutionary and he introduced the idea of sequence heterochrony (evolutionary changes in the sequence of developmental events). Despite the availability of detailed developmental data (e. g. Franz Keibel's 'Normal Tables'), Haeckel was unable to undertake a quantitative analysis of developmental data. This is now possible through computer-based analytical techniques such as event-pairing, which can extract important biological information from developmental sequences by mapping them onto established phylogenies. It may also yield data that can be used in phylogeny reconstruction, although the inherent 'non-independence' of the data may make this invalid. In future, the methods discussed here may be applied to the analysis of patterns of gene expression in embryos, or adapted to studying gene order on chromosomes.

Introduction

Embryonic development is a continuous process; the embryo possesses an ever-changing morphology. This continuum may be analysed by treating it as a series of discrete developmental events, where each event is the first

1431-7613/02/121/03–297/$ 15.00/0

appearance of a defined morphology, morphogenetic movement, or gene expression pattern. The order in which a given set of events takes place is known as a developmental sequence. Such sequences are of intrinsic interest because they define the ontogeny of an individual. Furthermore, they gain significance when viewed in an evolutionary context. Changes in sequence during evolution are known as sequence heterochronies, and have become increasingly studied as potential mechanisms of vertebrate evolution (Gould 1977; Gould 1982; Wray and Raff 1990; McKinney and McNamara 1991; Smith 2001).

Ernst Haeckel pioneered the study of developmental sequences. His approach was explicitly evolutionary, in that he attempted to account for the relationships among organisms (Richardson and Keuck in press). However, he did not use quantitative methods (and sometimes appears not to have used real data at all; Richardson 1995). Haeckel equated the developmental sequence with the phyletic series of presumed ancestors. Mismatches between developmental and phyletic sequences were caused, Haeckel said, by adaptations of the embryo to its environment. They could appear as additions, deletions, or heterochronies in the sequence (reviewed by Richardson and Keuck in press). Together, these mismatches were classified as 'caenogenetic' by Haeckel. Haeckel never analysed these heterochronies in detail. One reason for this is that the quantitative analysis of developmental sequences within a phylogenetic framework is impractical without the use of computers.

The aims and structure of this paper are as follows. First, we discuss the various sources of developmental data for single species and show how developmental sequences are derived from them. Second, we discuss the problems in comparing developmental sequences between species, which in turn hinder the analysis of developmental sequences in an evolutionary context. Third, current procedures for sequence analysis in an evolutionary context, including statistical methods and event-pairing are described. Finally, we conclude by considering future directions in this research field, such as the use of developmental sequence data in phylogenetic inference.

In our examples, we concentrate on the organogenetic period of vertebrates. This is because this period in vertebrates is relatively well studied; there is a long history of comparative embryological studies, including Haeckel's work on heterochrony (see Richardson and Keuck in press). However, the issues raised in this paper apply equally well to any organism at any time during its ontogeny where a developmental sequence can be observed.

1. The properties of developmentel data

Definition of 'events'

Developmental events may be regarded as a series of morphological states through which a given embryonic structure passes. For example, the development of the lens in humans may be divided into a series of events, such as the appearance of the lens placode, the start of the invagination of the placode, and the final separation of the placode from the surface ectoderm. Between each event, there is a transformation from one morphology to the next (see Figure 1 i).

The definition of events is clearly an arbitrary decision. However, well-defined events should reflect distinct changes in the embryo. Each event should be clearly defined, and unambiguously distinguishable from the transformation which precedes it. All events will possess a small, but finite duration. Events of long duration can be subdivided into one or more near-instantaneous events. For example, "closure of the neural tube" is a prolonged event in vertebrates; however, it can be redefined as three events of more limited duration: "first fusion of neural folds", "closure of the anterior neuropore", and "closure of the posterior neuropore".

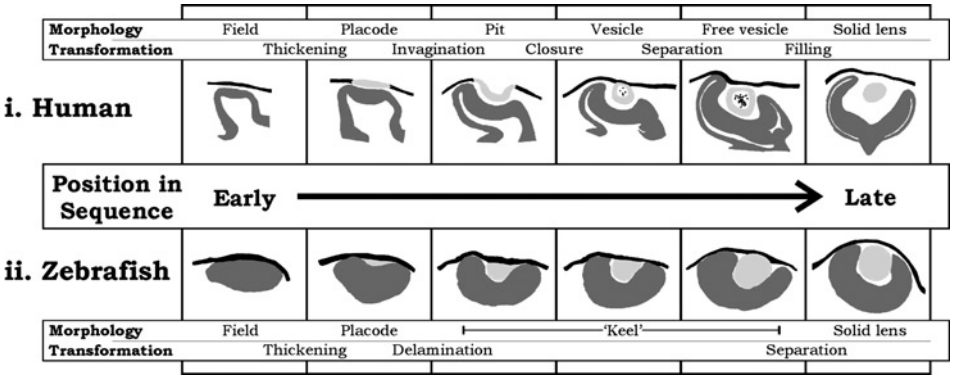

Fig. 1. The relationship between events and transformations in a developmental sequence. The formation of the lens (pale grey) from surface ectoderm (black) is shown schematically. The optic cup is shown in dark grey. Developmental 'events' are the first appearance of the morphology listed. There are transformations between each morphology. (**i**) The human lens is formed by the invagination of a thickened ectodermal placode. The placode forms a vesicle, which then separates from the surface ectoderm. The cavity of the vesicle reduces until a solid lens is formed. (**ii**) In the zebrafish, the lens is also formed from a thickened ectodermal placode. However, it does not invaginate, but forms a solid 'keel'. This separates from the surface ectoderm, to produce a solid lens. A vesicle is not formed. This raises the problem of homologising the process of lens formation in the two species (see Main Text).
(**i**) after O'Rahilly and Müller (1987), stages 13 to 17; (**ii**) after Li et al. (2000), stages 14-somite to Prim-25 (Kimmel et al. 1995).

Measuring the progress of development in individual species

Ideally, the development of a species should be studied by following an individual embryo *in vivo* through the entire course of its development (e. g. Chipman et al. 2000). In practice this is usually impossible, and data must be obtained from a collection of embryos fixed at different points in their development. Assembly of the developmental sequences in such cases thus requires some objective method to assess the relative progress of development in each embryo. This may be measured using references that are either external or internal to the dataset. We discuss each in turn.

External criteria: age and size

The most obvious external metrics of developmental progress are size and chronological age (e. g. crown-rump length in mammals, or estimated gestational ages). External criteria are frequently used in studies on laboratory animals (e. g. the mouse *Mus musculus*), where the time of fertilisation is known, and the conditions of development can be standardised. However, in many cases, age and size correlate poorly with morphological maturity (e. g. as noted for the Rhesus monkey, *Macaca mulatta*; Gribnau and Geijsberts 1981) due to variability in factors including incubation or gestation temperature, maternal nutrition, and genetic variation within a species (Hall and Miyake 1995). Also, in material collected from the wild, the chronological age is usually unknown.

Internal criteria: morphological staging

When external criteria are unavailable, or cannot be correlated with developmental progress, the events themselves are often used to divide developmental progress into developmental 'stages'. The term 'stage' can be used in a general sense to mean 'age' or 'maturity'. In this paper, however, we define stages more specifically as arbitrarily chosen, contiguous segments of a developmental sequence. Stages give a useful description of maturity, and circumvent the problems associated with using external references (Hall and Miyake 1995). The use of stages has a long history; Wilhelm His' staging tables for human development are a relatively early example (His 1880).

A single event may be used to define each stage. However, it is difficult to justify the selection of one event above any other to be the 'key' event on purely biological grounds. Instead, a broader summary of developmental progress can be achieved by using a cluster of events to define each stage (Hall and Miyake 1995). This also allows the developmental stage of an embryo to be assessed if one or more defining structures are damaged.

If abundant material is available, each stage can be based on observations of several embryos (e. g. Nieuwkoop and Faber 1994; O'Rahilly and Müller 2000). However, the decision on how many stages to define, and where

to place the boundaries between each stage, is arbitrary (Wheeler 1990). Often, the decision is dictated purely by the available material, with stages being based on single embryos.

The data: Developmental events

Literature sources and Normal Tables

Ideally, original comparative material should be collected to answer a specific research question. However, such material is often unavailable. It can be difficult to obtain a suitable range of material from species that have a low fecundity, or have a prolonged period of internal development (e. g. mammals, but also animals such as crocodiles that lay their eggs at an advanced stage of development; Ferguson 1985). Many species cannot be bred in captivity, and collecting material from the wild may be difficult, or might raise ethical and conservation issues. Comparative material exists in many museum collections (e. g. Richardson and Narraway 1999), although often only a limited range of material is available for each species.

Instead, suitable data can often be found in the literature, although the scope of these descriptions varies according to the material available, the research interests of the author, and the purpose of the description. For example, many descriptions are simply reference guides ('staging series'), devised to allow researchers to match a particular embryo easily and quickly against a standard description. They are often based only on external features (e. g. Hamburger and Hamilton 1951; Townsend and Stewart 1985; Kimmel et al. 1995). Other descriptions give detailed information on the development of a single organ system or time period, rather than an overview of the whole of development (e. g. Grandel and Merker 1998).

Keibel (1895) recognised that the majority of developmental descriptions then available were limited in detail. He therefore instituted a long-term project, extending the earlier work of His (1880) and Oppel (1891), to describe the development of a range of vertebrates, based on whole and serial-sectioned embryos. This project resulted in a 'benchmark' series of developmental descriptions published as standardised 'Normal Tables' (*Normentafeln*) or 'Normal Stages' (Keibel 1895; Keibel 1897–1938). Each table was produced by a specialist in the field, with Keibel acting as editor of the series (see Table 1 i). Studies by other authors also exist (see Table 1 ii).

Whatever the source of the developmental data, whether original observations or obtained from the literature, descriptions are usually not fully comprehensive. Thus, the descriptions are of varying suitability depending on the species of interest, and the questions asked. Below we discuss some other factors that must be considered when selecting developmental descriptions for use in studies of developmental sequences.

Table 1

i. Franz Keibel's *Normentafeln* (Normal Tables)

Higher Taxon	Species[1]	Common Name	Reference
Chondrichthyes			
Squaliformes: Squalidae	*Squalus acanthias*	Spiny Dogfish	Scammon 1911
Dipnoi			
Neoceratodontidae	*Neoceratodus forsteri* (*Ceratodus forsteri*)	Australian Lungfish	Semon 1901
Lepidosirenidae	*Lepidosiren paradoxa*	South American Lungfish	Kerr 1909
Lepidosirenidae	*Protopterus annectens*	African Lungfish	Kerr 1909
Lissamphibia			
Caudata: Salamandridae	*Triturus vulgaris* (*Molge vulgaris*)	Smooth Newt	Glaesner 1925
Caudata: Proteidae	*Necturus maculosus*	Mudpuppy	Eycleshymer and Wilson 1910
Caudata: Cryptobranchidae	*Andrias japonicus* (*Megalobatracus japonicus*)	Giant Salamander	Kudô 1938
Sauropsida			
Squamata: Lacertidae	*Lacerta agilis*	Sand Lizard	Peter 1904
Aves: Galliformes	*Gallus gallus* (*G. domestica*)	Domestic Chicken	Keibel and Abraham 1900
Aves: Charadriiformes	*Vanellus vanellus* (*V. cristatus*)	Lapwing	Grosser and Tandler 1909
Mammalia			
Rodentia: Muridae	*Rattus norvegicus*	Brown Rat	Henneberg 1937
Rodentia: Sciuridae	*Spermophilus citellus*	European Ground Squirrel	Völker 1922
Lagomorpha: Leporidae	*Oryctolagus cuniculus* (*Lepus cuniculus*)	Rabbit	Minot and Taylor 1905
Artiodactyla: Cervidae	*Capreolus capreolus* (*Cervus capreolus*)	Roe Deer	Sakurai 1906
Artiodactyla: Suidae	*Sus scrofa*	Domestic Pig	Keibel 1897
Primates: Tarsiidae	*Tarsius spectrum*	Spectral Tarsier	Hubrecht and Keibel 1907
Primates: Loridae	*Loris tardigradus* (*Nycticebus tardigradus*)	Slender Loris	Hubrecht and Keibel 1907
Primates: Hominidae	*Homo sapiens*	Human	Keibel and Elze 1908

Simultaneous events and the problem of resolution

Probably only a few developmental events occur at precisely the same time during ontogeny (i. e. are truly simultaneous; Smith 1997; Nunn and Smith 1998). Therefore, most inferences of simultaneity are artefactual

Table 1

ii. Selected Other Normal Tables[2]

Higher Taxon	Species[1]	Common Name	Reference
Actinopterygii			
Teleostei: Cyprinodontiformes	*Xiphophorus maculata* (*Platpoecilus maculata*)	Platyfish	Tavolga 1949
Lissamphibia			
Anura: Pipidae	*Xenopus laevis*	African Clawed Toad	Nieuwkoop and Faber 1994
Anura: Ranidae	*Rana temporaria* (*R. fusca*)	Common Frog	Kopsch 1952
Aves			
Aves: Galliformes	*Gallus gallus*	Domestic Chicken	Bellairs and Osmond 1998
Psittaciformes: Psittacidae	*Melopsittacus undulatus*	Budgerigar	Abraham 1901
Mammalia			
Didelphimorphia: Didelphidae	*Didelphis virginiana*	Virginia Opossum	McCrady 1938
Pholidota: Manidae[3]	*Manis javanica*	Malayan Pangolin	Huisman and de Lange 1937
Scandentia: Tupaiidae[3]	*Tupaia javanica*	Javan Tree Shrew	de Lange and Nierstrasz 1932
Rodentia: Caviidae	*Cavia porcellus*	Guinea Pig	Scott 1937
Rodentia: Muridae	*Mus musculus*	House Mouse	Theiler 1989
Primates: Hominidae	*Homo sapiens*	Human	O'Rahilly and Müller 1987

[1] Species name used in the table given in brackets.
[2] See Duellman and Trueb (1994: Table 5.6) for an extensive list of amphibian Normal Tables.
[3] Strictly speaking, these two papers are late additions to the Keibel 'Normal Tables'.

(Velhagen 1997): in attempting to subdivide the continuous process of ontogeny into discrete stages, some consecutive events will inevitably be forced into the same subdivision.

The "resolution" of a staged developmental series refers to our ability to observe a timing difference between two non-simultaneous events. Resolution depends both on the total number and distribution of stages described for a given developmental time period, and on the set of developmental events under investigation (Figure 2; see also Wheeler 1990). Ideally, there should be regular and small increments of maturation between each stage. A low resolution occurs when a sequence contains conspicuous gaps (i. e. extreme irregularities in the spacing of the stages or the distribution of the events under study). This is often a 'local' phenomenon, affecting only parts of the sequence.

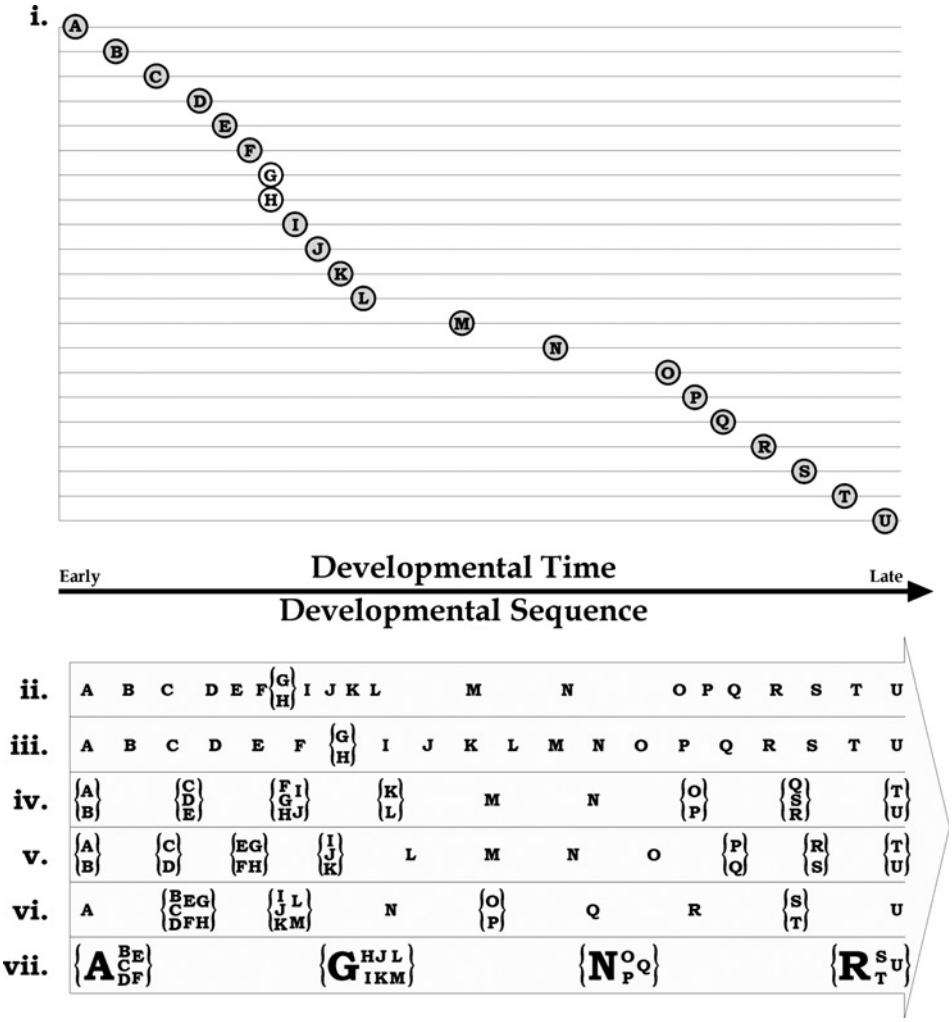

Fig. 2. The effect of stages on developmental sequences. (**i**) Abacus graph (Richardson 1999) showing the true absolute timing of 21 developmental events (**A**–**U**). The distribution of events is uneven, although only two events are simultaneous (**G** & **H**; highlighted in white). (**ii**–**vii**) Developmental sequences obtained from (**i**), with varying resolution. Events inferred to be simultaneous shown inside brackets. (**ii**) Sequence using absolute (chronological) time. This shows the order of events and their absolute timings. (**iii**) Ranked sequence. The order of events is retained, but their absolute timing relationships are lost. (**iv**) Staged sequence, based on 9 approximately evenly spaced observations. The division into stages has retained much of the information in (**ii**) and (**iii**), but has caused some artefactual inferences of simultaneity ('AIS'; e. g. **O** & **P**). (**v**) Staged sequence, based on 11 approximately evenly spaced observations. The greater number of stages has reduced the AIS. However, because the division between stages (observation points) varies between (**iv**) & (**v**), the AISs may be different (e. g. **O** & **P** and **P** & **Q**). Also note how in (**iv**) & (**v**) the uneven distribution of events results in local reductions in resolution (i. e. an uneven distribution of AISs). (**vi**) Staged sequence, based on 9 unevenly spaced observation points. Irregular stages can also lead to local reductions in resolution (e. g. **B**–**H**). (**vii**) Staged sequence, based on 4 'key' events (**A**, **G**, **N**, **R**). This results in 4 stages, as all other events are inferred to be simultaneous with one of the 'key' events. This sequence has a very low resolution.

However, even if the events are distributed evenly, gains in resolution follow a 'law of diminishing returns'. Beyond a certain point, using more stages will not increase resolution appreciably (Figure 2 iv, v).

Intraspecific variation

When a broad sample of embryos is available, small intraspecific variations may confound efforts to arrange them into a linear sequence of developmental maturity. A number of alternative sequences may be possible, depending on which events are inferred to have a variable timing. The Ontogenetic Sequence Analysis method ('OSA'; Colbert and Rowe 2001) can identify one or more most parsimonious sequence(s), from all the possible alternatives. A most parsimonious sequence ('MPS') is a sequence which infers the smallest set of variable events. If more than one MPS is identified, all will infer a set of equal size, although the members of the set will be different for each MPS. These MPSs can be treated in two ways. First, they may be treated as variant sequences and analysed separately. However, unless intraspecific variation is an area of interest, using variant sequences may prove an unnecessarily complication.

The second approach is to integrate the alternatives to produce a consensus sequence. Regions of the sequence common to all solutions may be broken into stages as described above. A region which varies between solutions can be treated as a single stage. This eliminates the variation by reducing the resolution of the consensus sequence in that region – all the potentially variable events are now perceived to occur simultaneously. Depending on the number of alternative solutions, and the number and type of events involved, such a local reduction of resolution may be preferable to using several variant sequences. For example, in the different embryos used to derive the Normal Table of the pig (Keibel 1897), the first contact of the neural folds variously occurs just before, during, or just after the formation of the first somite. However, neither event shows variation with respect to any other event, and no inductive relationship has been suggested between them. Thus they might safely be scored at a lower resolution and treated as simultaneous (but see 'Event-pairing', below).

A small degree of intraspecific variation is apparent (and occasionally discussed explicitly) in highly resolved Normal Tables. However, the precise extent of intraspecific variation in development is poorly known except for a few important studies (e. g. Mabee and Trendler 1996; Mabee et al. 2000 and references therein).

Assessing the quality of developmental data

What, then, makes for a good developmental series? Although it is easy to list of set of desirable properties, quantifying the quality of a developmental series is not as straightforward. In general, both low resolution and missing data are undesirable and should be avoided. However, there is no

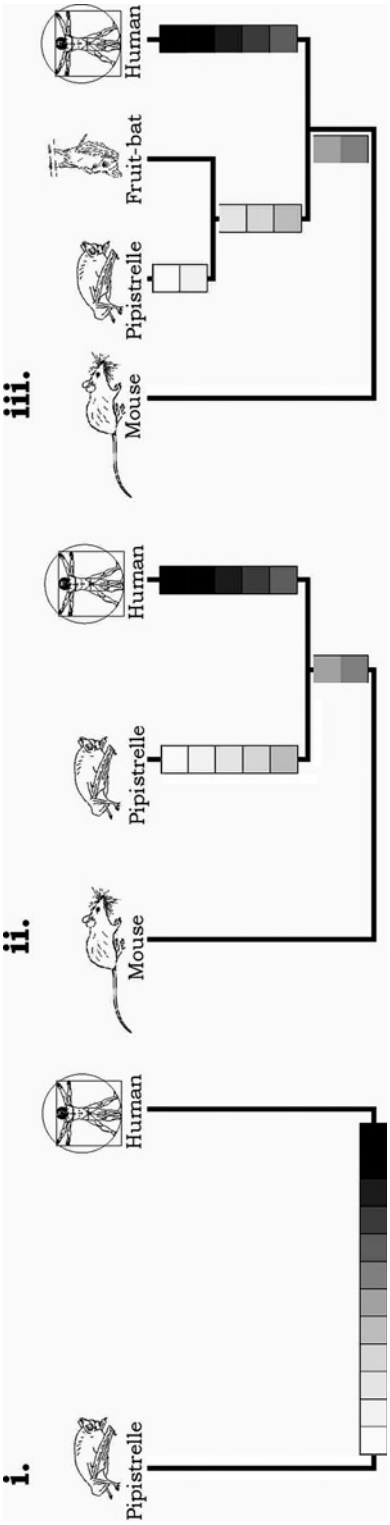

Fig. 3. The use of outgroups and sister groups in comparative studies. (i) The developmental sequences of a pipistrelle bat and a human are found to be different. However, it is unclear along which branch the differences arose, and therefore whether they might be correlated to the evolution of echolocation in bats. (ii) The inclusion of a mouse as an outgroup helps to highlight which changes are unique to the pipistrelle and which are unique to the human. (iii) Correlations between sequence changes and echolocation in the pipistrelle may be further investigated by the inclusion of a fruit bat, a sister group of the pipistrelle which does not echolocate.

hard-and-fast rule concerning how much either affects the suitability of a description for sequence analysis. Both resolution, and the amount of missing data, are easily quantified, but these values are only indirectly related to the usefulness of a description in a given analysis. In the end, investigators must make a value judgement on the desirability of including a particular event or species in the analysis, at the cost of introducing ambiguous data.

2. Phylogenetic framework: comparing development between species

A study of the evolution of development (for example, to identify heterochronies) requires comparisons between several species. In doing so, the importance of including an additional, outgroup species to infer the directionality of evolutionary change is well documented (Figure 3; Kitching et al. 1998). However, before such comparisons can be made, several points must be considered.

Homology of events

Phylogenetic analyses of developmental timing, as for standard morphological and molecular characters, require that the features being compared among species are

homologous (Smith 1997; Nunn and Smith 1998). In most cases, the homologies are clear. In other cases, the transformations between similar morphological states may not be the same. In other words, the same states arise through different processes. For instance, in the zebrafish (Figure 1 ii), two events in lens formation are shared with the human (Figure 1 i): the lens first appears as a thickened placode, and eventually becomes a solid lens separated from the surface ectoderm. However, the intervening events in the human (e. g. start of invagination of the placode, formation of vesicle) are not found in the zebrafish. Instead, the zebrafish placode delaminates as a solid mass and never forms a vesicle. Such instances should provoke an assessment of the assumptions of the primary homology of the states defining each event. In the above example, the lens placodes and the solid lenses are likely to share a common evolutionary history in both species; they are therefore homologous. Thus, the process of transformation between these two shared morphologies (and any different intervening events) must have been modified from the primitive condition in one or both species.

Similarly, serial homology can cause difficulties. For example, determining whether the tenth pair of post-otic somites is homologous between embryos of two different species can be difficult if the total somite count varies (Richardson et al. 1998). Unfortunately, questions involving serial homology are not easily answered at present, even in adult organisms (see Kitching et al. 1998), although the use of a molecular perspective may help provide another valuable line of evidence (Lauder 1994).

Inclusion of simultaneous events

Event-simultaneity deserves careful consideration because the artefactual simultaneity of two events may cause errors when reconstructing developmental sequence evolution. For example, consider events **O** and **P** in the developmental sequences in Figure 2. The timing difference is resolved in some sequences (ii, iii, v), but are apparently simultaneous in others (iv, vi, vii). This gives the appearance of a real difference between the two sets of developmental sequences. If they were compared to one another, an evolutionary change would be inferred (i. e., with the events changing from being not simultaneous to simultaneous or vice versa).

For these reasons, Velhagen (1997) suggested that simultaneous events be treated as missing data unless evidence for true simultaneity exists. However, this highly conservative approach may not be appropriate in most cases, as the errors caused by trying to accommodate missing data (e. g. Platnick et al. 1991) may be worse than those caused by possible artefactual simultaneity. At the very least, data on simultaneous events should be recorded, and their effects investigated.

Comparing measurements of developmental progress

Chronological age

As described above, measurements of chronological age rarely correlate well with developmental maturity in a single species. When comparing development between species, this problem is compounded by the fact that the total developmental timespan differs widely, even in closely related organisms. Therefore, if chronological age is to be used as the metric of developmental maturity, comparisons between species require elaborate 'normalisation' of the timing data (e. g. Dettlaff and Dettlaff 1961). This, in turn, requires homologous starting and end-points for calibration.

Fertilisation forms an obvious starting point from which to measure development, but using it in this way is beset by practical problems. As noted above, for material collected in the wild, the time of mating is rarely known. Further, in animals with internal fertilisation, specialisations such as delayed implantation, arrested development, or sperm storage disrupt the direct connection between mating, fertilisation and development.

More critically, there does not appear to be an obvious, homologous end-point to development. In humans, the end of the embryonic period is variously, and arbitrarily, defined as the time at which the lateral palatine processes fuse, the humerus begins to ossify (Streeter 1949), when 90% of the over 4 500 named adult structures have been laid down (O'Rahilly 1979), 3 lunar months (Hamilton et al. 1946) or 12 weeks (Langman 1975) after the onset of the last menstruation. These end-points all occur before birth. In other taxa, such as teleost fish or amphibians, embryonic development is described as continuing after hatching (Nieuwkoop and Faber 1994; Kimmel et al. 1995). Clearly then, birth or hatching may involve different life history stages in different taxa. This is true even within a group such as mammals, in which marsupials are born at a much earlier morphological stage than eutherians (Beard 1896).

Key events and morphological staging

Stages might be expected to form more useful points of comparison because they are independent of chronological age. Haeckel (1910) and Witschi (1956) attempted to define a series of universal embryonic stages for the vertebrates (e. g. blastula, gastrula and so on), each based on a common key character. However, this process is difficult to justify on biological grounds, for the same reasons that apply to species-level key-events (see above and cf. Ballard 1964). Furthermore, defining universal stages assumes at least some biological reality or importance to developmental stages. We would argue that this is invalid given our contention that stages are an arbitrary subdivision of the continuum of development.

In practical terms, the need for 'universality' means that each key event has to involve a structure seen in the entire group. Thus, the events have to be the first appearance of symplesiomorphic vertebrate structures (nerve chord, pharyngeal slits, etc.), which leads to a small number of stages of very long duration (Figure 2 vii). Most importantly, however, the existence of heterochrony means that any single key event can shift its timing relative to other events or vice versa (Richardson 1995). Thus, two species at the same key stage may have very different morphologies beyond shared possession of the key event. The end of each stage has to be defined by the appearance of the next key character (and thus the start of the next stage). It is therefore possible that heterochrony could move two 'universal' key characters to be simultaneous (apparently obliterating a stage), or even reverse their sequence of appearance (giving apparently homologous stages non-homologous endpoints).

Although they provide a better summary of developmental maturity, staging systems based on clusters of events are still subject to the same problems. As with the 'key' events, the division between stages is arbitrary, and not necessarily comparable between species (cf. Figure 2 iii–vi). However, even when stages are defined with specific comparisons in mind, heterochronic changes mean that a given cluster of events may not occur at the same time in the species under study. For example, in lungfishes, Kerr (1909) found it impossible to correlate his stages for *Lepidosiren paradoxa* and *Protopterus annectens* with those described by Semon (1901) for *Neoceratodus forsteri*.

3. Methods of analysing sequence data

In this section, we review some of the major methods developed for analysing developmental sequences in an evolutionary context. We give the most attention to the method of 'event-pairing', which we hold to show the greatest potential.

Since neither chronological timing nor developmental stages can easily or rigorously be compared across species, most methods use the relative position or ranks of events in the developmental sequence. Because of the inherent loss of information this entails (cf. Figure 2 i, ii), sequence heterochrony will be visualised only when events change their positions relative to one another. Events which change their absolute developmental timing, but not their relative position in the sequence will be undetected.

Statistical methods for comparing sequences

These methods are united by ranking the events in each developmental sequence (e.g. 1st, 2nd, 3rd …). The ranked data can then be analysed in two

ways. The first compares entire developmental sequences using rank correlation measures such as Spearman's rank correlation (Mabee and Trendler 1996) or Kendall's coefficient of concordance (Nunn and Smith 1998). The second compares the mean ranks of individual events using ANOVA (Nunn and Smith 1998).

Both methods of analysis share a number of shortcomings. First, they only provide overall (phenetic) measures of similarity. They are therefore not suitable for analysis within a phylogenetic framework (Mabee and Trendler 1996), although computer simulation can produce a null distribution that partly accounts for phylogenetic relatedness (e. g., Nunn and Smith 1998).

Second, the procedures are limited to detecting differences between only two groups (as in Figure 3 i) because including outgroup information is difficult. By themselves, the methods cannot localise where any changes have occurred, nor which group shows the primitive or derived pattern (see Jeffrey et al. 2002 b).

Third is the problem of missing data. Rank correlation measures cannot accommodate missing entries at all, while the remaining methods require normalisation of the ranks to correct for the effects of missing data (and therefore missing ranks). In effect, these problems mean that only species with timing information for all the events under study can be included in the analysis. In practice, this can severely limit the number of species or events that can be included.

Heterochrony plots

A heterochrony plot (Schlosser 2001) compares a set of events common to the developmental sequence of two species. The position of each event on the plot is determined by its time of occurrence in each species (Figure 4). The timing data can be relative (e. g. ranks) or absolute (chronological). If the events show the same order and similar intervals between them in both species, the heterochrony plot will be linear. The slope of the line reflects any difference in the developmental timespan of the two species. Events that show heterochrony are highlighted because they fall above or below this line. Schlosser (2001) used this method to investigate modularity in development; that is, whether certain clusters of events form functional or developmental units, and thus tend to shift position in unison.

Although they are a useful visual aid for studying aspects of developmental timing, heterochrony plots have some limitations. First, the incorporation of phylogenetic information is incomplete or awkward. Because there is no procedure for inferring the conditions at the internal nodes of a tree, heterochrony plots are only maximally informative when sister species are compared. Furthermore, outgroup (or non-sister species) comparisons can only be made by plotting each of the ingroup species in turn against the

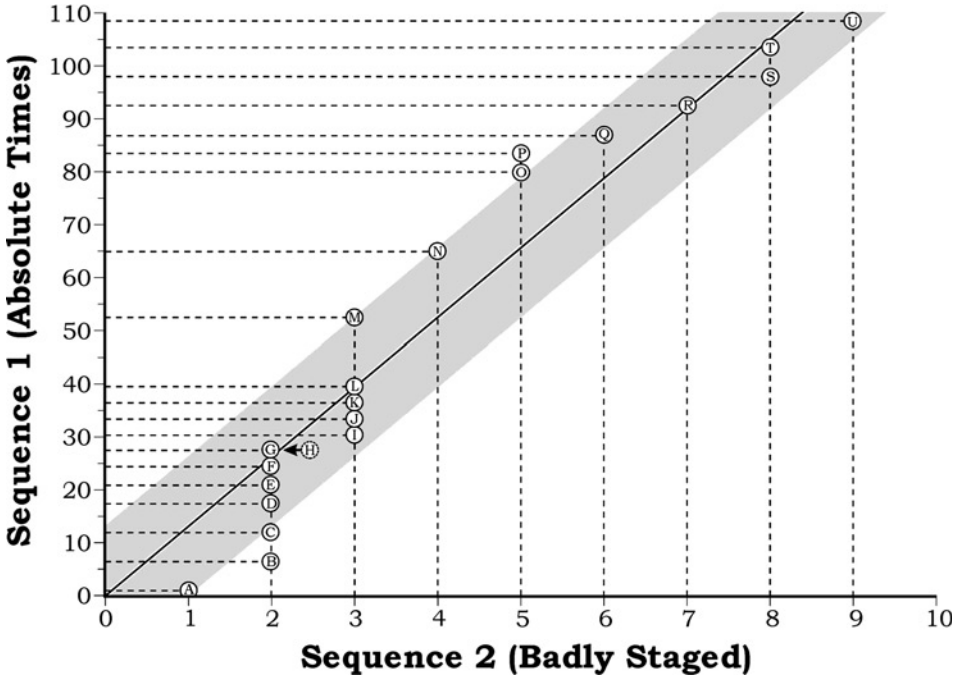

Fig. 4. A heterochrony plot (Schlosser 2001). In this hypothetical example, two developmental sequences of 21 events (**A**–**U**) each are plotted against one another. Events **G** & **H** occur simultaneously in both sequences – for clarity event **H** is shown to one side (indicated by a dashed-outline and arrow). A line-of-best-fit passing through the origin (black line) shows the relationship between the two sequences. Events falling above or below this line are inferred to have a different timing in the two sequences. 'Confidence intervals' around this line are given by the stage intervals for each sequence (broad grey band for the x-axis; narrow white band for the y-axis). However, before hypotheses of heterochrony can be formulated, care must be taken to assure that the intervals between the stages are regular in each sequence. Although this plot appears to reveal heterochrony, the data for both sequences are obtained from the same 21 events in Figure 2; Sequence 1 from Figure 2 ii and Sequence 2 from Figure 2 vi.

outgroup. Second, the choice of line clearly influences the interpretation of which events deviate from it, and thus the inference of which, if any, modules have shifted their position. However, there is no objective criterion for choosing the line.

A third, more fundamental, limitation is that heterochrony plots cannot distinguish between (and are confounded by) different forms of heterochrony. The plots assume that there is a comparable rate of development in both species. However, one or both species might have an uneven rate of development (e. g. alternating periods of slow and rapid development) or the spacing of their stages might be irregular. On the resulting heterochrony plot, the events will not form a straight line, but describe a curve or irregular line. This might not be apparent, and the non-linearity could mistakenly be identified as sequence heterochrony (i. e. events sitting

above or below an inappropriate best-fit line; Figure 4). Schlosser (2001) advocated the use of 'sampling thresholds', indicating the maximum precision of the staged developmental data. This would reduce the problem, but not eliminate it.

Event-pairing

In our opinion, the most promising method for comparing developmental sequences is that of event-pairing, which was developed in the 1990s (Mabee and Trendler 1996; Smith 1996; Velhagen 1997). It has been successfully used to investigate heterochrony within various vertebrate groups such as therian mammals (Smith 1996; Smith 1997), thamnophiine snakes (Velhagen and Savitzky 1998), and amniotes (Jeffery et al. 2002 a).

The procedure

Event-pairing operates by reducing the developmental sequence of each species to statements about the relative timing of pairs of events. Two events in a developmental sequence (**A** and **B**) may have only one of three timing relationships: **A** may occur before **B**, **A** and **B** may occur simultaneously, or **A** may occur after **B**. These relationships may be represented by simple numerical scores (for example, 0, 1 and 2, respectively). Missing data can also be accommodated in event-pair analyses: if the timing of an event is unknown in a particular species, any event-pairs involving that event are scored as missing ('?'). For each species, the relative timing of every possible pair of events from the developmental sequence is recorded, giving $^1/_2(N^2–N)$ event-pairs for N events.

Because the event-pairs comprise events that are homologous amongst the species studied, the data can be compared directly – they can, in fact, be used as characters in standard phylogenetic procedures. For example, if the data are mapped onto a phylogenetic tree, the evolution of the relative timing of each pair of events may be reconstructed, and ancestral states estimated. Unlike the statistical methods, heterochronic changes can be localised to specific branches on the phylogenetic tree using this procedure (see Jeffery et al. 2002 a, b).

Accommodating intraspecific variation

Event-pairing allows information about intraspecific timing variation (i. e. variant sequences) to be included in the analysis, a highly desirable feature. This is accomplished through the use of polymorphism coding in many parsimony programs (e. g. *PAUP**; Swofford 2001), which allows a single cell in the character matrix to have multiple values. In the example of the pig (see 'Intraspecific Variation', above), if variation in the timing of neural fold contact causes it to arise just before, during, or just after the forma-

tion of the first somite, their event-pair will be {012}. This creates the potential for a single consensus sequence to retain information on variable events, without the need for local reductions in resolution. Polymorphism coding will be the most informative when the intraspecific variation excludes one possible timing relationship (i. e. **A** occurs before or simultaneously with **B**, but never after it).

Polymorphic characters are routinely excluded from phylogenetic analyses, ostensibly because they are felt to be less reliable or informative (cf. Wiens 1995). However, as discussed above, events that display variation in their timing may do so relative only to adjacent events in the sequence. Within a single species, the use of polymorphic coding will account for any intraspecific variation in adjacent events, while including potentially informative data about those events' constant timing relationship with other, more distant, events.

Relative timing changes to polarised shifts

When event-paired data are mapped onto a phylogenetic tree, patterns of evolution become apparent. Shared derived changes in character state (synapomorphies) along various branches of the tree represent a shared change in the timing relationship between two events. However, without further analysis, little can be said about the actual heterochronies that gave rise to the timing change (i. e. which event(s) actually moved). For example, if the event-pair comparison of events **A** and **B** changes from "**A** is before **B**" to "**A** is after **B**", it gives no information whether one or both events shifted, nor in which direction they shifted (see Jeffery et al. 2002 a, b).

Elsewhere (Jeffery et al. 2002 b), we present a method of overcoming this problem by analysing the event-pair synapomorphies along a branch *en bloc*. This method of 'event-pair cracking' operates on the principle that events which are actively moving in development will change their timing relationship: i) relative to many other (non-moving) events, and ii) do so in a consistent manner. In the above example where **A** and **B** have changed their timing relationship, an examination of the other event-pair synapomorphies along the same branch might indicate that event **A** also changed its timing relationship with other events (say **C**, **D**, **E** and **F**) whereas **B** only moves relative to **A**. Thus the most parsimonious explanation is that only **A** has shifted its position in the developmental sequence, moving relative to five static events (**B**, **C**, **D**, **E** and **F**). This could be examined further by checking if **C**, **D**, **E** and **F** have changed their relative timing with any events other than **A** (e. g. have **C** and **D** changed their relative timing?). A final check is examining whether all the changes involving event **A** are 'coherent' – that is, do all the event-pair changes suggest that **A** is moving in the same direction, either earlier or later in the sequence? Collating such data from all the synapomorphies along a given branch

(using standardised procedures) often suggests that the majority of event-pair changes can be accounted for by a small subset of events shifting early or later in the developmental sequence (e. g., see Jeffrey et al. 2002 a, b).

4. Future uses of event-paired data

We conclude by examining some possible extensions of the event-pair method in studying developmental timing changes. Largely, this concerns the possibility of constructing phylogenetic trees using the evolutionary signal that is apparent in heterochrony data.

Phylogenetic inference

Inferring a phylogeny using parsimony assumes that the character data are independent (Smith 1994); non-independence among characters may bias the phylogenetic estimate. However, the analysis of developmental sequences via event-pairing involves two forms of non-independence. One arises from the nature of the developmental sequence itself ("ontogenetic non-independence"), whilst the other is a consequence of the event-pairing coding procedure ("coding non-independence"; Smith 1997). Neither form of non-independence is detrimental when heterochrony data are mapped onto an existing phylogeny. However, both are problematic when heterochrony data are used to infer a phylogeny.

Ontogenetic non-independence

Alberch (1985) first drew the important distinction between ontogenetically dependent ("causal") and independent ("temporal") events in a developmental sequence. Dependence in a sequence occurs when one event is obligatory for the development of another for functional reasons. This will constrain sequence variation to some degree. An example is the lens of the eye, which cannot appear before the lens placode has formed (Figure 1). Inductive interactions between tissues (e. g. those leading to the formation of the lens placode; reviewed by Baker and Bronner-Fraser 2001) can also constrain the sequence of events.

Some degree of ontogenetic non-independence is probably present in all morphological data sets (Lovejoy et al. 1999). This may reflect the hierarchical nature of development – ultimately all adult structures are derived from a single fertilised egg cell. In phylogenetic inference, its effect can be to increase the weight of some evolutionary changes. For example, if ontogenetic non-independence constrains ten events to shift position in unison, they represent (in evolutionary terms) a single character describing a single evolutionary change. However, phylogenetic analysis assumes that all ten events are free to shift independently, giving this single evolutionary

change ten times more 'weight' (influence on the outcome) than it should possess.

Coding non-independence

Event-pairing adds further non-independence because each event is compared to every other event in the sequence. Therefore, in a sequence of N events, each event will be involved in N-1 event-pairs. Thus, when an event shifts its position in the developmental sequence, it can alter the scores of

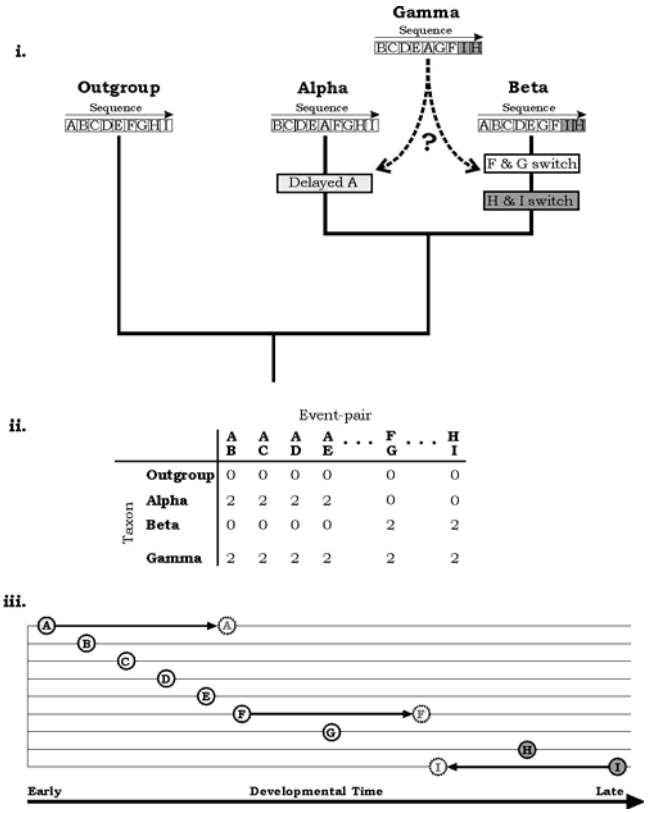

Fig. 5. Hypothetical example showing the effect of non-independence on a parsimony analysis of event-paired data. (**i**) The test species, Gamma, possesses both of the timing shifts found in Alpha (delayed event **A**) and Beta (events **F** & **G** switch places; events **H** & **I** switch places). Gamma should cluster more parsimoniously with Beta (with which it shares two changes) rather than Alpha (with which it shares only one change). (**ii**) Event-pairs derived from the sequences. The delay in event **A** affects more event-pairs ("characters") than the **F** & **G** and **H** & **I** shifts combined, thereby causing Gamma to cluster more parsimoniously with Alpha. (**iii**) Abacus graph showing how the three sequences changes seen in (**i**) could in fact be the result of three heterochronic shifts of equal absolute magnitude. The number of event-pairs affected depends on the distribution of surrounding events. This distribution is highly unpredictable, affected both by ontogeny and which events were selected for analysis.

Only informative event-pairs are displayed in (**ii**) for clarity; all other event-pair combinations were constant (state 0) across species.

anything from one event-pair (if it swaps place with its neighbour) to N-1 event-pairs (if it shifts from the beginning of the sequence to the end).
This 'coding non-independence' introduces a potentially severe bias into a parsimony analysis. A larger timing shift will tend to produce changes in more event-pairs than will a smaller shift, even though they both represent single instances of heterochrony. However, because of the assumption of independence, the topology of the inferred tree will tend to be influenced more strongly by larger shifts (Figure 5 i, ii). This behaviour is appropriate if it is assumed that larger timing shifts should carry more weight in a phylogenetic analysis (e. g., because they are more unlikely and therefore rarer). However, there is little biological evidence that the evolutionary cost of a shift is directly proportional to its magnitude. Indeed, the number of event-pairs affected by a sequence heterochrony can be purely a result of the distribution of the events in the sequence (see Figure 5 iii). Thus, if the phylogenetically relevant information is the existence of the shift, and not its magnitude, some form of correction is needed.
Despite the inherent non-independence of the event-pairing method, parsimony analysis of event-paired developmental data has been shown to give reasonable results for thamnophiine snakes (Velhagen, 1997) and the major tetrapod lineages (Jeffrey et al. 2002 a). This suggests that the non-independence, while present and theoretically problematic, may not be that detrimental in practice. It is likely that, in the complex changes seen in real data sets, a simple conflict between one large and several small shifts seldom arises. Instead, the overlap of timing shifts will tend to break up any large shifts. We are currently using computer simulations to test the behaviour of event-paired data when inferring phylogeny.

Molecular developmental data

At present, studies of developmental sequences in general (and the use of the event-pairing method in particular) have been restricted to morphological data. This is understandable, as morphological events are easier to observe, and a range of developmental descriptions is available in the literature. However, the sequence of expression of developmental genes could be studied using the same phylogenetic methods reviewed above. There have been many studies on the causal relationships between gene expression and morphology (for example, the role of Hox genes in vertebrate body-segment identity). Recent discoveries have underlined the importance of the relative timing of such expression patterns (e. g. Tabin and Johnson 2001). Again, the phylogenetic methods discussed in this paper provide a means to examine the relationship between changes in gene expression and morphology, possibly by testing whether the events constitute a developmental module or not.

Conclusions

Haeckel's pioneering work on developmental sequences demonstrated the occurrence of heterochrony in vertebrate evolution. Developmental descriptions such as Keibel's Normal Tables provided a database of developmental information for many vertebrate species. These works are a useful source of information, although the quality of the data (i.e. data-coverage and resolution of the sequence) must be considered.

Sequence heterochronies could not be studied quantitatively until the development of computer-based techniques in the latter part of the 20th century. Data sets compiled to analyse sequence heterochrony are complex; this is caused in part by the lack of an absolute developmental timeframe. Another problem is resolution of the observations, and the related phenomenon of artefactual simultaneity.

Several methods now exist to analyse developmental sequences in a phylogenetic context. All have limitations, although we believe that event-pairing is the most flexible. The key to this method is that it is able to account for phylogenetic relatedness and thus to distinguish between uninformative shared primitive and informative shared derived similarities. Using this method, it may finally be possible to use developmental sequence information to elucidate both evolutionary history and the constraints of developmental patterning.

Acknowledgements

This paper is based on ideas presented at the "Haeckel and Modern Biology" Symposium at the 6th International Congress of Vertebrate Morphology, held at Jena in July 2001. OBE was supported by an NSERC postdoctoral fellowship and by the van der Leeuw Fonds; JEJ was supported by the Leverhulme Trust. We thank Paula Mabee and Kathleen Smith for constructive comments on the ideas in this paper.

References

Abraham, K. (1901) Beiträge zur Entwicklungsgeschichte des Wellensittichs (*Melopsittacus undulatus*). Anat. Hefte **17:** 589–669.

Alberch, P. (1985) Problems with the interpretation of developmental sequences. Syst. Zool. **34:** 46–58.

Baker, C. V. H., Bronner-Fraser, M. (2001) Vertebrate cranial placodes. I. Embryonic induction. Dev. Biol. **232:** 1–61.

Ballard, W. W. (1964) Comparative anatomy and embryology, Ronald Press, New York.

Beard, J. (1896) On certain problems of vertebrate embryology. Gustav Fischer. Jena.

Bellairs, R., Osmond, M. (1998) The atlas of chick development. Academic Press. San Diego.

Chipman, A. D., Haas A., Tchernov E., Khaner O. (2000) Variation in anuran embryogenesis: differences in sequence and timing of early developmental events. J. Exp. Zool. **288:** 352–365.
Colbert, M. W., Rowe, T. (2001) Ontogenetic sequence analysis: using parsimony to characterize developmental hirearchies. J. Morphol. **248:** 218.
de Lange, D., Nierstrasz, H. F. (1932) Tabellarische Übersicht der Entwicklung von *Tupaia javanica* Horsf. Oosthoek Verlag. Utrecht.
Dettlaff, T. A., Dettlaff, A. A. (1961) On relative dimensionless characteristics of the development duration in embryology. Arch. Biol. (Liege) **72:** 1–16.
Duellman, W. E., Trueb, L. (1994) Biology of amphibians. Johns Hopkins University Press. Baltimore.
Eycleshymer, A. C., Wilson, J. A. (1910) Normal plates of the development of *Necturus maculosus*. Gustav Fischer Verlag. Jena.
Ferguson, M. W. J. (1985) Reproductive biology and embryology of crocodilians. In: Gans, C., Billett, F. (eds) Biology of the Reptilia. Wiley. New York. pp. 329–491.
Glaesner, L. (1925) Normentafeln zur Entwicklungsgeschichte des gemeinen Wassermolchs (*Molge vulgaris*). Gustav Fischer Verlag. Jena.
Gould, S. J. (1977) Ontogeny and phylogeny. Belknap Press. Cambridge, Massachusetts.
Gould, S. J. (1982) Change in developmental timing as a mechanism of macroevolution. In: Bonner, J. T. (ed.) Evolution and development. Springer. Berlin. pp. 333–346.
Grandel, H., Merker, S. S. (1998) The development of the paired fins in the zebrafish (*Danio rerio*). Mech. Devel. **79:** 99–120.
Gribnau, A. A. M., Geijsberts, L. G. M. (1981) Developmental stages in the Rhesus monkey (*Macaca mulatta*). Springer. Berlin.
Grosser, O., Tandler, J. (1909) Normentafel zur Entwicklungsgeschichte des Kiebitzes (*Vanellus cristatus* Meyer). Gustav Fischer. Jena.
Haeckel, E. (1910) The evolution of man. Watts & Co. London.
Hall, B. K., Miyake, T. (1995) How do embryos measure time? In: McNamara, K. J. (ed.) Evolutionary Change and Heterochrony. Wiley. New York. pp. 3–20.
Hamburger, V., Hamilton, H. L. (1951) A series of normal stages in the development of the chick embryo. J. Morphol. **88:** 49–92.
Henneberg, B. (1937) Normentafeln zur Entwicklungsgeschichte der Wanderratte (*Rattus norvegicus* Erxleben). Gustav Fischer. Jena.
His, W. (1880) Die Anatomie menschlicher Embryonen. Vogel. Leipzig.
Hubrecht, A. A., Keibel, F. (1907) Normentafeln zur Entwicklungsgeschichte des Koboldmaki (*Tarsius spectrum*) und des Plumplori (*Nycticebus tardigradus*). Gustav Fischer. Jena.
Huisman, F. J., de Lange, D. (1937) Tabellarische Übersicht der Entwicklung von *Manis javanica* Desm. Oosthoek Verlag. Utrecht.
Jeffery, J. E., Bininda-Emonds, O. R. P., Coates, M. I., Richardson, M. K. (2002 a) Analyzing evolutionary patterns in amniote embryonic development. Evol. Dev.
Jeffery, J. E., Richardson, M. K., Coates, M. I., Bininda-Emonds, O. R. P. (2002 b). Analyzing developmental sequences within a phylogenetic framework. Syst. Biol.
Keibel, F. (1895) Normentafel zur Entwicklungsgeschichte der Wirbeltiere. Anat. Anz. **11:** 255–234.
Keibel, F. (1897) Normentafel zur Entwicklungsgeschichte des Schweines (*Sus scrofa domesticus*). Gustav Fischer. Jena.
Keibel, F. (1897–1938) Normentafeln zur Entwicklungsgeschichte der Wirbelthiere. Gustav Fischer. Jena.
Keibel, F., Abraham, K. (1900) Normentafel zur Entwicklungsgeschichte des Huhnes (*Gallus domesticus*). Gustav Fischer. Jena.
Keibel, F., Elze, C. (1908) Normentafel zur Entwicklungsgeschichte des Menschen. Gustav Fischer. Jena.
Kerr, J. G. (1909) Normal plates of the development of *Lepidosiren paradoxa* and *Protopterus annectens*. Gustav Fischer. Jena.
Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., Schilling, T. F. (1995) Stages of embryonic development of the zebrafish. Dev. Dynam. **203:** 253–310.
Kitching, I. J., Forey, P. L., Humphries, C. J., Williams, D. M. (1998) Cladistics (2e). Oxford University Press. Oxford.

Kopsch, F. R. (1952) Die Entwicklung des braunen Grasfrosches *Rana fusca* Roesel. Georg Thieme Verlag. Stuttgart.
Kudô, T. (1938) Normentafeln zur Entwicklungsgeschichte des japanischen Riesensalamanders (*Megalobatrachus japonicus* Temminck). Gustav Fischer. Jena.
Lauder, G. V. (1994) Homology, form, and function. In: Hall, B. K. (ed.) Homology: the hierarchical basis of comparative biology. Academic Press. San Diego. pp. 151–196.
Li, Z., Joseph, N. M., Easter, S. S. (2000) The morphogenesis of the zebrafish eye, including a fate map of the optic vesicle. Dev. Dynam. **218:** 175–188.
Lovejoy, C. O., Cohn, M. J., White, T. D. (1999) Morphological analysis of the mammalian postcranium: a developmental perspective. Proc. Natl. Acad. Sci. USA **96:** 13247–13252.
Mabee, P. M., Olmstead, K. L., Cubbage, C. C. (2000) An experimental study of intraspecific variation, developmental timing, and heterochrony in fishes. Evolution **54:** 2091–2106.
Mabee, P. M., Trendler, T. A. (1996) Development of the cranium and paired fins in *Betta splendens* (Teleosti: Percomporpha): intraspecific variation and interspecific comparisons. J. Morphol. **227:** 249–287.
McCrady, E. (1938) Embryology of the opossum. Amer. Anat. Mem. **16:** 1–233.
McKinney, M. L., McNamara, K. J. (1991) Heterochrony, the evolution of ontogeny. Plenum. New York.
Minot, C. S., Taylor, E. (1905) Normal plates of the development of the rabbit (*Lepus cuniculus* L.). Gustav Fischer. Jena.
Nieuwkoop, P. D., Faber, J. (1994) Normal Table of *Xenopus laevis* (Daudin) (2e). Garland Publishing. London.
Nunn, C. L., Smith, K. K. (1998) Statistical analyses of developmental sequences: the craniofacial region in marsupial and placental mammals. Amer. Nat. **152:** 82–101.
Oppel, A. (1891) Vergleichung des Entwicklungsgrades der Organe. Fischer. Jena.
O'Rahilly, R. (1979) Early human development and the chief sources of information on staged human embryos. Europ. J. Obstet. Gynec. Reprod. Biol. **9:** 273–280.
O'Rahilly, R., Müller, F. (1987) Developmental stages in human embryos. Carnegie Institution. Washington.
O'Rahilly, R., Müller, F. (2000) Prenatal ages and stages – measures and errors. Teratology **61:** 382–384.
Peter, K. (1904) Normentafel zur Entwicklungsgeschichte der Zauneidechse (*Lacerta agilis*). Gustav Fischer. Jena.
Platnick, N. I., Griswold, C. E., Coddington, J. A. (1991) On missing entries in cladistic analysis. Cladistics **7:** 337–343.
Richardson, M. K. (1995) Heterochrony and the phylotypic period. Dev. Biol. **172:** 412–421.
Richardson, M. K. (1999) Vertebrate evolution: the developmental origins of adult variation. BioEssays **21:** 604–613.
Richardson, M. K., Allen, S. P., Wright, G. M., Raynaud, A., Hanken, J. (1998) Somite number and vertebrate evolution. Development **125:** 151–160.
Richardson, M. K., Keuck, G. (in press) Haeckel's ABC of evolution and development. Biol. Rev.
Richardson, M. K., Narraway, J. (1999) A treasure house of comparative embryology. Int. J. Dev. Biol. **43:** 591–602.
Sakurai, T. (1906) Normentafel zur Entwicklungsgeschichte des Rehes (*Cervus capreolus*). Gustav Fischer. Jena.
Scammon, R. E. (1911) Normal plates of the development of *Squalus acnathias*. Gustav Fischer. Jena.
Schlosser, G. (2001) Using heterochrony plots to detect the dissociated coevolution of characters. J. Exp. Zool. (Mol. Dev. Evol.) **291:** 282–304.
Scott, J. P. (1937) The embryology of the guinea pig. Amer. J. Anat. **60:** 397–432.
Semon, R. (1901) Normentafeln zur Entwicklungsgeschichte des *Ceratodus forsteri*. Gustav Fischer. Jena.
Smith, A. B. (1994) Systematics and the fossil record. Blackwell Scientific. London.
Smith, K. K. (1996) Integration of craniofacial structures during development in mammals. Amer. Zool. **36:** 70–79.
Smith, K. K. (1997) Comparative patterns of craniofacial development in eutherian and metatherian mammals. Evolution **51:** 1663–1678.

Smith, K. K. (2001) Heterochrony revisited: the evolution of developmental sequences. Biol. J. Linn. Soc. **73:** 169–186.
Streeter, G. L. (1949) Developmental horizons in human embryology (fourth issue). A review of the histogenesis of cartilage and bone. Contrib. Embryol. **33:** 149–169.
Swofford, D. L. (2001) PAUP*. Phylogenetic analysis using parsimony (*and other methods) (v.4). Sinauer Associates. Sunderland, Massachusetts.
Tabin, C. J., Johnson, R. L. (2001) Clocks and Hox. Nature **412:** 780–781.
Tavolga, W. N. (1949) Embryonic development of the Platyfish (*Platpoecilus*), the swordtail (*Xiphophorus*) and their hybrids. Bull. AMNH **94:** 165–229.
Theiler, K. (1989) The house mouse: atlas of embryonic development. Springer. New York.
Townsend, D. S., Stewart, M. M. (1985) Direct development in *Eleutherodactylus coqui* (Anura: Leptodactylidae): a staging table. Copeia **1985:** 423–436.
Velhagen, W. A. (1997) Analyzing developmental sequences using sequence units. Syst. Biol. **46:** 204–210.
Velhagen, W. A., Savitzky, A. H. (1998) Evolution of embryonic growth in thamnophiine snakes. Copeia **1998:** 549–558.
Völker, O. (1922) Normentafel zur Entwicklungsgeschichte des Ziesels (*Spermophilus citillus*). Gustav Fischer. Jena.
Wheeler, Q. D. (1990) Ontogeny and character phylogeny. Cladistics **6:** 225–268.
Wiens, J. J. (1995) Polymorphic characters in phylogenetic systematics. Syst. Biol. **44:** 482–500.
Witschi, E. (1956) Development of vertebrates. Saunders. Philadelphia.
Wray, G. A., Raff, R. A. (1990) Pattern and process heterochronies in the early development of sea urchins. Semin. Dev. Biol. **1:** 245–251.