Rare genomic changes as a tool for phylogenetics

Antonis Rokas and Peter W.H. Holland

DNA sequence data have offered valuable insights into the relationships between living organisms. However, most phylogenetic analyses of DNA sequences rely primarily on single nucleotide substitutions, which might not be perfect phylogenetic markers. Rare genomic changes (RGCs), such as intron indels, retroposon integrations, signature sequences, mitochondrial and chloroplast gene order changes, gene duplications and genetic code changes, provide a suite of complementary markers with enormous potential for molecular systematics. Recent exploitation of RGCs has already started to yield exciting phylogenetic information.

Antonis Rokas is at the Institute of Cell, Animal and Population Biology, Ashworth Laboratories, King's Buildings, West Mains Road, University of Edinburgh, Edinburgh, UK EH9 3JT (a.rokas@ed.ac.uk); Peter Holland is at the School of Animal and Microbial Sciences, University of Reading, Whiteknights, PO Box 228, Reading, UK RG6 6AJ (p.w.h.holland@reading.ac.uk).

In recent years, considerable progress has been made in the field of molecular phylogenetics. A significant driving force has been the increasing technical ease of DNA sequencing, which has led to the dominance of primary sequence data as indicators of the historical relationships between taxa. Important advances have also occurred in the computational analysis of DNA sequence data[1], such as improved methods for modelling patterns of nucleotide substitutions[2]. However, the task of phylogenetic reconstruction using molecular sequences is not without problems. To a large extent, these stem from the fact that the dominant methods for molecular phylogeny reconstruction exploit nucleotide substitutions (plus, in some cases, single-site insertions or deletions) as indicators of divergence or common descent. Convergent evolution of nucleotide bases, differing substitution rates among sites and lineages[3], saturation of mutations at variable sites[4], nonindependent substitutions among sites[5] and functional constraints at the molecular level[6] are just a small sample of the potential caveats that apply when using these types of data. As a result, phylogenetic hypotheses based on primary sequence data can sometimes be equivocal[7,8], whereas others can simply be incorrect[9,10]. The advent of the genomic era has brought the opportunity to consider other types of information embedded in DNA sequences. Here, we consider the phylogenetic use of large-scale mutations – rare genomic changes (RGCs; Box 1), which occur relatively infrequently. Researchers have already started using RGCs for inferring relationships between living organisms.

Rare genomic changes

We define RGCs as large-scale mutational changes that have occurred in the genomes of particular clades. Examples of RGCs (Table 1) include intron indels, retroposon integrations, signature sequences, changes in organelle gene order, gene duplications and genetic code variants. Most RGCs represent changes caused by single (or a few) mutational events; in our discussion of RGCs we do not include genomic characteristics that are, most probably, the end result of multiple processes (e.g. genomic compositional contrasts[11]). Until recently, many studies mapped RGCs onto existing phylogenies to gain insight into their mode of evolution. The consensus that has emerged is that RGCs are often evolutionarily conserved and phylogenetically informative. We believe the time has come to turn the question around: what can RGCs tell us about phylogenies themselves?

RGCs provide an independent source of phylogenetic information, largely immune from some of the problems that affect primary sequence data. A major difficulty with this approach is the identification of these rare mutations in the clades of interest. However, the increasing automation of molecular techniques has brought us to the dawn of the genomic era where tremendous amounts of information, freely available in the primary literature and public databases, are generated. Additionally, protocols have been developed for the targeted identification of many, if not all, RGCs. Here, we argue that the application of RGCs to phylogenetics can offer new insights into evolutionary history. Furthermore, in cases where primary sequence data generate conflicting or equivocal results, RGCs offer an independent way of evaluating alternative phylogenetic scenarios.

RGCs as 'Hennigian' markers

The field of phylogenetics has been strongly influenced by the founder of the cladistic methodology, the German entomologist Willi Hennig. Hennig argued that only shared derived characters (synapomorphies; Box 1) should be used as indicators of common descent. Plotting the distribution of synapomorphies is the essence of cladistic reconstruction. The principal hindrance to this task is homoplasy (see Box 1 for definition). In general, character states that arise rarely will not be prone to extensive convergent or parallel evolution, which should contribute to a low level of homoplasy. Although the precise frequency of occurrence of most RGCs has not been robustly estimated, large-scale mutations are generally rare. Additionally, precise secondary loss of the character (homoplasy because of reversion) is likely to be extremely rare for most large-scale mutations and has been demonstrated to be so in some cases [e.g. short interspersed element insertions (SINEs)[12]; Box 1]. Therefore, with respect to homoplasy, RGCs might constitute good markers of common descent. In Table 1, we provide a summary of the characteristics and phylogenetic applicability of various categories of RGC, and in the next section we expand on a few examples published recently to demonstrate their potential use.

 0169-5347/00/$ – see front matter PII: S0169-5347(00)01967-4

Of fish and flies: intron indels as clade markers

The power and robustness of RGCs is well demonstrated by the study of Venkatesh *et al.*[13], in which intron indels (Box 1) were used to investigate fish phylogeny. Venkatesh *et al.* identified seven intron positions (in five genes) that are present in the pufferfish *Takifugu rubripes* but not in the homologous genes of mammals. Four introns were also found in the rhodopsin gene that were present in the ancestral chordate rhodopsin gene (as inferred by their presence in basal chordates, such as lampreys and skates, and in the more apical lineage of mammals) but were absent in the pufferfish. Several ray-finned fish species (class Actinopterygii) were screened for the presence or absence of these eleven intron indels, and these data were used to reconstruct evolutionary relationships[13]. Only one indel showed considerable homoplasy and an unclear phylogenetic signal; all the others were unique synapomorphies able to resolve phylogenetic relationships. It is noteworthy that some of the relationships resolved, such as the basal position of bichirs (*Polypterus* spp.) within the Actinopterygii, have proved contentious using primary sequence comparisons.

Another recent use of an intron indel as a phylogenetic character deals with the placement of the insect order Strepsiptera within holometabolous insects[14]. Strepsipteran forewings resemble the hindwing balancing organs of flies (order Diptera), which are known as halteres. Among other phylogenetic scenarios, an affinity of Strepsiptera to Coleoptera has been widely discussed, based primarily on the use of hindwings for flight in both orders. An alternative proposal is a sister group relationship with Diptera[15,16]. In this case, halteres could be homologous, but a radical homeotic mutation might have reversed their position in Strepsiptera[15]. Evidence from morphology is equivocal[17] and 18S rDNA sequence data have generated a lively debate between researchers favouring different phylogenetic reconstruction methods[8,16]. Rokas *et al.* noted a unique intron insertion in the homeobox of the *engrailed* gene of Diptera and Lepidoptera, which is absent from other insects and all outgroups[14]. Possession of the intron in Strepsiptera would support a sister group relationship with Diptera, whereas its absence would argue against this affinity. Cloning of the Strepsipteran homologue of *engrailed* showed that the intron is absent in Strepsiptera, thus suggesting that the halteres of Strepsiptera and Diptera might not represent a rare case of natural homeotic transformation but a remarkable case of convergent evolution[14] (Fig. 1).

Box 1. Glossary

Bilateria: the bilaterally symmetrical animals.
Homoplasy: a general term denoting that the acquisition of the same character state in two taxa is not because of common descent. This can arise by parallel evolution (independent acquisition from the same ancestral state), convergent evolution (independent acquisition from different ancestral conditions) or secondary loss (reversion from the derived to the ancestral condition).
Indel: an insertion or deletion event.
LINEs (long interspersed elements): a class of retroposons that are capable of self-transposition.
Orthology: the relationship between two homologous loci derived from a speciation event.
Paralogy: the relationship between two homologous loci derived from a duplication event.
Polyphyly: when a group does not include the most recent common ancestor of all its members.
Protein domain: a well defined region within the protein. It can be distinguished on the basis of function or structure. For example, the homeodomain is a 60-amino acid domain shared by proteins encoded by homeobox genes.
Protein motif: any stretch of contiguous sequence within a protein that has been evolutionarily conserved.
Rare genomic change (RGC): a large-scale mutational change that has occurred in the genome of a particular clade.
Retroposons: the class of transposable elements that relocate in the genome via an RNA intermediate using the enzyme reverse transcriptase.
Signature sequences: shared conserved insertions or deletions in proteins or RNAs.
SINEs (short interspersed elements): a particular class of retroposons that have lost the ability to transpose themselves (to transpose they use another class of mobile elements, LINEs).
Synapomorphy: a shared derived character state that suggests a monophyletic grouping.

Of SINEs and LINEs

Retroposons (Box 1) belong to the group of transposable elements that use an RNA-mediated mode of transposition[12]. Retroposon integrations, especially from the class of SINEs (retroposons that lack the ability for self-amplification), have been used successfully as phylogenetic markers; an application pioneered by Okada and colleagues in the 1990s (Refs 18,19). It has been argued that SINE integrations come close to being 'perfect' markers of common descent because integration is apparently random and irreversible, and because most eukaryotic genomes have an abundance of SINE elements[12]. Their presence or primitive absence can also be readily detected by PCR amplification across integration sites. Successful applications of SINEs include the generation of convincing

Table 1. Summary of useful characteristics of rare genomic change (RGC) markers for phylogenetic purposes[a]

Marker	Taxonomic resolution	Homoplasy	Taxa in which RGCs are applicable
Intron indels	Wide ranging	Low	Eukaryotes
Retroposons (SINEs and LINEs)[b]	Within orders	Zero to very low	Animals
Signature sequences	Wide ranging	Unknown but recognizable	All life
mtDNA genetic code variants	Phyla to classes	Low to moderate	Eukaryotes
Nuclear DNA genetic code variants	Phyla	Low to moderate	All life
mtDNA gene order	Wide ranging (phyla to families)	Low to moderate in animals. High in plants, fungi and protists	Eukaryotes
cpDNA gene order	Families	Low	Plants
Gene duplications	Wide ranging	Unknown	All life
Comparative cytogenetics	Within phyla	Unknown	All life (lateral gene transfer is prevalent in prokaryotes)

[a]For more detailed information, see references cited in the text.
[b]Abbreviations: SINEs, short interspersed elements; LINEs, long interspersed elements.

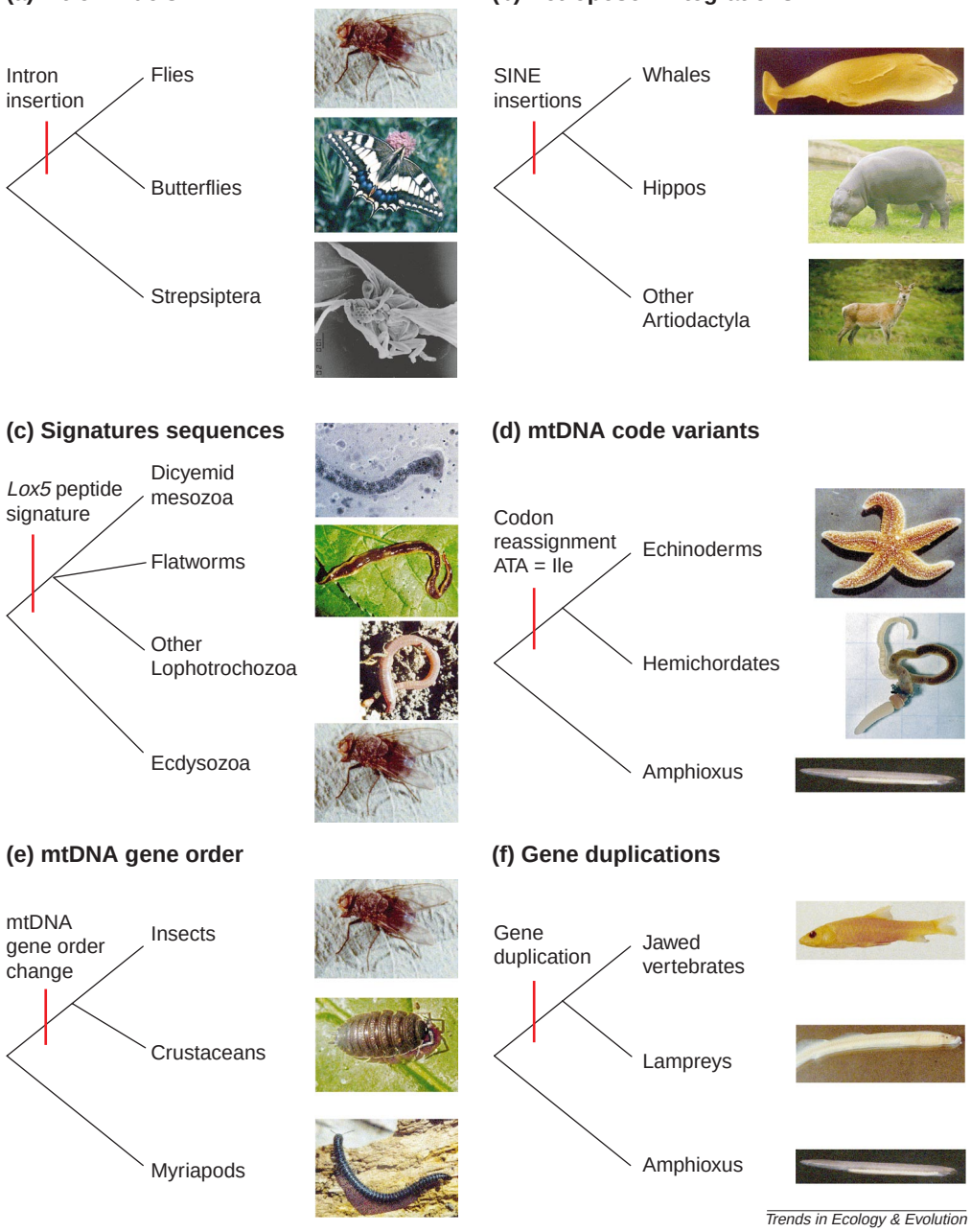

Fig. 1. Examples of rare genomic changes (RGCs) as phylogenetic markers. (a) An intron insertion in the gene *engrailed* suggests that Strepsiptera are not a sister group to Diptera (flies)[14]. (b) Retroposon integrations using short interspersed elements (SINEs) have established the sister group relationship of whales and hippos, to the exclusion of other Artiodactyla[20]. (c) Hox gene signature sequences have robustly supported the division of protostome invertebrates into the lophotrochozoan and ecdysozoan clades[25], and have identified dicyemid mesozoa as lophotrochozoans[26]. (d) A codon reassignment in the mitochondrial genetic code suggests that echinoderms and hemichordates are sister groups[34], a result supported by sequence data[35]. (e) mtDNA gene order in mitochondrial genomes supports the common grouping of insects and crustaceans, with myriapods as an outgroup[39]. (f) Repeated events of gene duplication have occurred in the lineage leading to vertebrates[46]. *Photographs reproduced, with permission, from G. Brown (fly/whale/flatworm/earthworm/crustacean/myriapod/starfish/fish/lamprey), S. Ferguson (hippo), P. Holland (butterfly), A. Rokas (Strepsiptera), M. Kobayashi (mesozoa), D. Remsen and the Marine Biological Laboratory (hemichordate) and J. Pemberton (deer).*

support for a sister group relationship between whales and hippopotamuses[20], also known as the 'whippo' hypothesis (Fig. 1), and detailed insight into salmonid fish phylogeny[19]. Criticisms of the use of SINEs include the nonindependence of SINE insertions[21] (several can be integrated at the same time, although at different sites), incomplete lineage sorting[21,22] (although this applies to all characters), the considerable amount of work needed for their development[12] and practical limits to detection beyond ~30% difference in sequences flanking orthologous elements[12] (Box 1). In our view, these are simply factors to be considered when designing or interpreting phylogenetic studies; they do not detract significantly from the robustness of SINE markers. Another class of retroposons are long interspersed elements (LINEs; Box 1); the main difference from SINEs being their ability for self-amplification. LINEs have been used not only for determining the cladogenetic pattern but also for dating speciation events. Verneau *et al.*[23] exploited the fact that LINE elements belonging to the L1 family rapidly generate defective copies, which are retained in the genome and mutate at the neutral rate, to resolve and date the phylogenetic history of the rodent genus *Rattus*.

Animals, archaebacteria and archezoa: the use of signature sequences

The complementary use of primary sequence data and RGCs for phylogenetic purposes is shown by attempts to reconstruct the interphyletic relationships of animals. Recent studies using 18S rDNA sequences have suggested a three-branched Bilateria (Box 1) tree comprising the Deuterostomia, the Lophotrochozoa and the Ecdysozoa[3]. Lophotrochozoans include spiral cleaving phyla, such as molluscs, annelids, platyhelminths and nemerteans, plus the lophophorates; whereas the Ecdysozoa include arthropods, onychophorans, priapulids and nematodes (all of which moult). This proposal was controversial because it demanded a radical restructuring of the classic tree of animal phyla. It would also mean abandoning several well known super phyletic groupings, such as Acoelomata (animals without a coelom, which are traditionally basal in the animal tree) and Articulata (segmented protostomes), and implies that some apparently 'simple' animals, such as flatworms, are actually highly degenerate. Radical hypotheses often require independent support before they are accepted; in this case, complementary supporting data have come from RGCs within the Hox gene clusters[24,25]. The proteins

encoded by many Hox genes possess specific sequence motifs near the homeodomain, which is a domain common to all Hox genes. These sequence motifs have helped distinguish orthologous and paralogous (Box 1) Hox genes. Each of the three major clades has its own unique Hox genes that do not have identifiable orthologues in the others. In other words, gene duplications have yielded distinct genes in each lineage and these have acquired unique signature protein motifs (Box 1). For example, the lophotrochozoans share *Lox2*, *Lox4*, *Lox5*, *Post-1* and *Post-2*, whereas the ecdysozoans share *Ubx* and *Abd-B* (Ref. 25).

As well as providing independent support for the controversial Lophotrochozoa and Ecdysozoa clades, this approach has been used to investigate the affinities of a particularly enigmatic animal phylum: the dicyemid mesozoa[26]. These are microscopic parasites of squid and octopus, with an amazingly simple body plan consisting of a solitary axial cell surrounded by a single layer of 10–40 ciliated outer cells. Morphology and 18S rDNA sequence data have previously failed to adequately resolve their phylogenetic position. Recently, Kobayashi *et al.*[26] cloned the *Lox5* gene from a dicyemid, including the diagnostic *Lox5* peptide, thus demonstrating that these animals are almost certainly highly degenerate members of the Lophotrochozoa clade (Fig. 1). Indeed, dicyemids represent one of the most extreme cases of secondary simplification of morphology known in the animal kingdom.

The Hox gene data represent a special case of 'signature sequences' (Box 1). In Hox genes, the existence of distinct signature sequences in different genes, and in different clades, suggests that these motifs (such as the *Lox5* peptide) have biochemical functions. In other examples, insertions might have little functional significance; nonetheless, they can be used as RGCs for phylogenetic reconstruction; for example, there is an ongoing debate in prokaryote phylogenetics about whether archaebacteria are monophyletic. Recently, the paraphyly hypothesis of archaebacteria has been supported by several signature sequences[27,28]. However, it should be noted that historical associations within prokaryotes are still incompletely resolved owing to extensive subsequent gene transfer. No molecular marker is immune from this all-pervading complication[29]; for example, a signature sequence in the gene *hsp70* used to support the paraphyly of archaebacteria[28] is also present in one of the three copies of the gene in *Escherichia coli*, suggesting a possible recent transfer[4]. Other important studies using the signature-sequence approach include a confirmation that the archezoa are true eukaryotes that have lost mitochondria[30], and an investigation of branchiopod crustacean phylogeny[31]. In branchiopod crustaceans, three unique helices in 18S rDNA were used to distinguish cladocerans from other branchiopods, demonstrating that useful sequences can be found in RNA, as well as in protein sequences.

Deviant codes and shuffled genes

Several organisms use genetic codes that deviate from the standard 'universal' code. These 'deviant' codes can be useful markers for higher level phylogenetics. Keeling and Doolittle[32] showed that a genetic code in which TAA and TAG codons encode glutamine, rather than termination, is used by almost all diplomonads, with the exception of the genus *Giardia*, which employs the standard genetic code. This argues for an early divergence of *Giardia* in the evolution of diplomonads and is in agreement with phylogenies from primary sequence data[32]. The diplomonad deviant code has also been found in certain green algae and in ciliates, showing that homoplasious changes can occur. Mitochondrial genomes have the widest variety of deviant codes, whereas plastids show no deviation from the universal (so far)[32,33]. Variant mtDNA codes in animals have been studied in some detail, aided by complete sequences of mtDNA from a wide range of animals. For example, a sister group relationship between echinoderms and hemichordates is supported by the assignment of the ATA codon to the amino acid isoleucine[34], as well as by sequence analyses[35] (Fig. 1), although the same reassignment has occurred independently in Cnidaria.

Gene order changes, particularly in circular genomes such as mitochondria and chloroplasts, comprise another type of RGC that has already proved useful in phylogenetics[36]. These arrangements, effected by inversions, translocations and duplications, generally affect several adjacent genes. They are unlikely to be reversed precisely because of their complexity; therefore, they satisfy one of the principal criteria demanded of the perfect phylogenetic marker. The second criterion, low levels of homoplasy, is also predicted to be true because convergence or parallelism would imply bias towards particular gene rearrangements or gene orders. Isolated cases of convergence have been detected[37], suggesting bias in some taxa; however, this does not seem to be a widespread problem. Some key phylogenetic problems have been tackled using mtDNA order as a marker, with definitive results. For example, until recently it was widely accepted that insects and myriapods were close relatives within the arthropods; indeed, these two primarily terrestrial taxa share many derived morphological characters. Several lines of evidence, including developmental data and primary sequence comparisons, have challenged this relationship, raising the alternative possibility of a crustacean–insect clade[38]. This suggestion is effectively confirmed by the shared presence of a rare tRNA translocation within insects and crustacean mtDNA, which is not seen in myriapods, chelicerates, tardigrades, onychophorans or outgroups[39] (Fig. 1). In most animal taxa, changes to mtDNA gene order are rare, making these markers useful for higher level phylogenetics[39,40]; although one exception might be the gastropod molluscs, where mtDNA gene order is extremely variable[41]. Similarly, plant, fungi and protist mtDNAs display rapid genome reshaping, making gene order a more appropriate marker for lower-level phylogenetics[42].

Chloroplast DNA (cpDNA) gene order has been exploited in a similar way to mtDNA gene order. For example, in 1987 Jansen and Palmer used a cpDNA inversion within the sunflower family to propose the basal position of the Barnadesiinae, with implications for biogeography and morphological evolution in this group[43]. More recently, Doyle *et al.*[44] surveyed 132 legume genera for the occurrence of a 50 kb inversion, finding evidence that at least two tribes within the legumes were polyphyletic. A qualitatively different sort of rearrangement from those discussed above is deletion. For example, monophyly of the conifers is supported by loss of one copy of an inverted repeat found in cpDNA (Ref. 45).

Other potential RGCs

The list of RGCs we have described so far is not exhaustive; several other categories of large-scale mutation exist, some of which have potential for phylogenetics. For example, gene duplications have not yet been widely exploited. One difficulty is technical: unless a family of genes is arranged in a tandem array, discerning whether a duplicated copy of a gene exists is difficult because absence of

evidence does not equate with evidence for absence. This problem cannot be definitively overcome except by the acquisition of complete genome sequences. Until this becomes faster, easier and cheaper, gene duplications represent a potentially untapped source of markers.

Gene families that have proved most amenable for tracing gene duplications include Hox genes and globin genes. It is no coincidence that these form stereotyped clustered arrangements that permit extra genes to be readily cloned. In the case of Hox genes, duplication of the entire gene cluster is deduced to have occurred on the vertebrate lineage, after divergence from the cephalochordate (amphioxus) lineage[46]. This dispels the view that cephalochordates are degenerate vertebrates (Fig. 1), although in reality this notion has had few supporters in the past century. There is also good evidence that the Hox gene clusters underwent additional duplications somewhere within the ray-finned fish lineage; however, more taxa need to be surveyed before this event can be used as a phylogenetic marker[47]. Although gene duplications are sufficiently widespread to be used as phylogenetic markers, there is still the potential for homoplasy. If new genes can be exploited for new roles (or to refine old roles), convergent duplication and retention is an ever present possibility. Somewhat paradoxically, gene losses might prove more useful markers than gene duplications. Although homoplasy is still a real possibility, at least reversion is virtually impossible. For example, after an additional round of Hox gene cluster duplications in ray-finned fish, approximately 21 individual gene losses (plus one cluster loss) must have occurred in the lineage leading to zebrafish (*Danio rerio*)[47]. This large number of independent events provides great scope for refinement of ray-finned fish phylogeny.

The study of the differences in chromosome structure and appearance between species has given rise to the field of comparative cytogenetics[48,49]: another source of RGCs that are potentially useful for phylogenetics. By comparing chromosomes, a phylogeny can be constructed based on the minimum number of rearrangements required or the maximum number of shared segments. Existing data are limited and come mainly from mammals, but there are exciting prospects[49]. As partial and complete genome sequences are obtained from an ever-growing number of species, the resolution of this approach can be greatly refined. Inversions, translocations and duplications, at the scale of one to a few genes, have occurred extensively in eukaryotic nuclear genomes and should provide a plethora of phylogenetic markers in the future.

A concluding mix of caution and optimism

One obstacle that makes some researchers feel uneasy about the use of RGCs is the absence of statistical evaluation[4,22]. This concern stems mainly from analogy with primarily sequence comparisons. Understanding the forces that shape sequence evolution is a necessary prerequisite to using sequence data for phylogenetics and for evaluating the statistical robustness of trees. To reach the same degree of sophistication in the analysis of RGCs demands greater knowledge about the mechanisms that generate RGCs because this will affect their rate of production, character independence, mutational biases and reversibility. Some of these parameters are reasonably well understood for some RGCs (notably SINE insertions and gene losses), but there is much more to be learnt. Nonetheless, the usefulness of several categories of RGC has been tested by comparison with phylogenies inferred by other methods (morphological and molecular). With few exceptions, RGCs have performed exceptionally well. Therefore, we feel that 'psychological constraints'[4] about statistical evaluation should be put aside, while applicability and robustness are tested further. Additionally, we stress that the 'Hennigian' framework is not the only one that can be employed when attempting to reconstruct a phylogeny based on RGCs. A statistical approach is also possible (e.g. maximum likelihood and Bayesian analysis) and will surely be of help, especially as more 'messy' data sets are obtained.

It is an inescapable (if uncomfortable) fact that a few good characters might contain more phylogenetic 'truth' than many poor ones. We do not suggest that all RGCs are necessarily 'good' markers; we certainly do not propose that they are a panacea for phylogenetics. Indeed, we have already alluded to cases of convergence, parallelism and reversion. However, we do believe that the suite of characters that we refer to as RGCs harbours enormous potential. They have already contributed some robust insights into important phylogenetic debates, such as the origin of whales, arthropod relationships, deuterostome phylogeny and diversification of the protostome invertebrates. Each of these 'new' phylogenies has wider implications, not only for evolutionary biology but also for biogeography, developmental biology and other areas. In a time when the acquisition of molecular data is outpacing analysis, it is worth recalling Darwin's comment[50]: '...we possess no pedigrees or armorial bearings; and we have to discover and trace the many diverging lines of descent in our natural genealogies, by any characters which have long been inherited.'

Acknowledgements

A.R. acknowledges partial funding from a NERC studentship; research in P.W.H.H.'s laboratory is funded by BBSRC. We thank G. Stone for his tolerance, and N. Okada and two anonymous referees for comments on this article. Special thanks to G. Brown, S. Ferguson, M. Kobayashi, J. Pemberton, and D. Remsen and the Marine Biological Laboratory for providing photographs. P. Preston very kindly allowed the use of specimens from the Natural History collections of the University of Edinburgh.

References

1 Hillis, D.M. *et al.*, eds (1996) *Molecular Systematics*, Sinauer
2 Yang, Z. *et al.* (1994) Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* 11, 316–324
3 Aguinaldo, A.M.A. *et al.* (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387, 489–493
4 Philippe, H. and Laurent, J. (1998) How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* 8, 616–623
5 Averof, M. *et al.* (2000) Evidence for a high frequency of simultaneous double-nucleotide substitutions. *Science* 287, 1283–1286
6 Lee, M.S.Y. (1999) Molecular phylogenies become functional. *Trends Ecol. Evol.* 14, 177–178
7 Forterre, P. and Philippe, H. (1999) Where is the root of the universal tree of life? *Bioessays* 21, 871–879
8 Huelsenbeck, J.P. (1998) Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biol.* 47, 519–537
9 Naylor, G.J.P. and Brown, W.M. (1998) Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* 47, 61–76
10 Philippe, H. (1997) Rodent monophyly: pitfalls of molecular phylogenies. *J. Mol. Evol.* 45, 712–715
11 Campbell, A. *et al.* (1999) Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9184–9189
12 Shedlock, A.M. and Okada, N. (2000) SINE insertions: powerful tools for molecular systematics. *Bioessays* 22, 148–160
13 Venkatesh, B. *et al.* (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10267–10271

14 Rokas, A. *et al.* (1999) Intron insertion as a phylogenetic character: the *engrailed* homeobox of Strepsiptera does not indicate affinity with Diptera. *Insect Mol. Biol.* 8, 527–530

15 Whiting, M.F. and Wheeler, W.C. (1994) Insect homeotic transformation. *Nature* 368, 696

16 Whiting, M.F. *et al.* (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from the 18S and 28S ribosomal sequences and morphology. *Syst. Biol.* 46, 1–68

17 Kristensen, N.P. (1999) Phylogeny of endopterygote insects, the most successful lineage of living organisms. *Eur. J. Entomol.* 96, 237–253

18 Kido, Y. *et al.* (1991) Shaping and reshaping of salmonid genomes by amplification of transfer RNA-derived retroposons during evolution. *Proc. Natl. Acad. Sci. U. S. A.* 88, 2326–2330

19 Murata, S. *et al.* (1993) Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6995–6999

20 Nikaido, M. *et al.* (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10261–10266

21 Miyamoto, M.M. (1999) Perfect SINEs of evolutionary history? *Curr. Biol.* 9, R816–R819

22 Hillis, D.M. (1999) SINEs of the perfect character. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9979–9981

23 Verneau, O. *et al.* (1998) Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. *Proc. Natl. Acad. Sci. U. S. A.* 95, 11284–11289

24 Balavoine, G. (1998) Are platyhelminthes coelomates without a coelom? An argument based on the evolution of Hox genes. *Am. Zool.* 38, 843–858

25 de Rosa, R. *et al.* (1999) Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399, 772–776

26 Kobayashi, M. *et al.* (1999) Dicyemids are higher animals. *Nature* 401, 762

27 Rivera, M.C. and Lake, J.A. (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257, 74–76

28 Gupta, R.S. (1998) What are archaebacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Mol. Microbiol.* 29, 695–707

29 Doolittle, W.F. (1999) Lateral genomics. *Trends Biochem. Sci.* 24, M5–M8

30 Hashimoto, T. *et al.* (1998) Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6860–6865

31 Crease, T.J. and Taylor, D.J. (1998) The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of branchiopod crustaceans. *Mol. Biol. Evol.* 15, 1430–1446

32 Keeling, P.J. and Doolittle, W.M. (1997) Widespread and ancient distribution of a noncanonical genetic code in diplomonads. *Mol. Biol. Evol.* 14, 895–901

33 Palmer, J.D. and Delwiche, C.F. (1998) The origin and evolution of plastids and their genomes. In *Molecular Systematics of Plants II: DNA Sequencing* (Soltis, D.E. *et al.*, eds), pp. 375–409, Kluwer

34 Castresana, J. *et al.* (1998) The mitochondrial genome of the hemichordate *Balanoglossus carnosus* and the evolution of deuterostome mitochondria. *Genetics* 150, 1115–1123

35 Bromham, L.D. and Degnan, B.M. (1999) Hemichordates and deuterostome evolution: robust molecular phylogenetic support for a hemichordate 1 echinoderm clade. *Evol. Dev.* 1, 166–171

36 Boore, J.L. (1999) Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780

37 Mindell, D.P. *et al.* (1998) Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10693–10697

38 Averof, M. and Akam, M. (1995) Insect–crustacean relationships: insights from comparative developmental and molecular studies. *Philos. Trans. R. Soc. London Ser. B* 347, 293–303

39 Boore, J.L. *et al.* (1998) Gene translocation links insects and crustaceans. *Nature* 393, 667–668

40 Boore, J.L. *et al.* (1999) Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* (amphioxus). *Mol. Biol. Evol.* 16, 410–418

41 Kurabayashi, A. and Ueshima, R. (2000) Complete sequence of the mitochondrial DNA of the primitive opisthobranch gastropod *Pupa strigosa*: systematic implication of the genome organization. *Mol. Biol. Evol.* 17, 266–277

42 Lang, B.F. *et al.* (1999) Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397

43 Jansen, R.K. and Palmer, J.D. (1987) A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Acad. Sci. U. S. A.* 84, 5818–5822

44 Doyle, J.J. *et al.* (1996) The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol. Phylog. Evol.* 5, 429–438

45 Raubeson, L.A. and Jansen, J.K. (1992) A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochem. Syst. Ecol.* 20, 17–24

46 Garcia-Fernàndez, J. and Holland, P.W.H. (1994) Archetypal organization of the amphioxus Hox gene cluster. *Nature* 370, 563–566

47 Stellwag, E.J. (1999) Hox gene duplication in fish. *Semin. Cell Dev. Biol.* 10, 531–540

48 O'Brien, S.J. and Stanyon, R. (1999) Ancestral primate viewed. *Nature* 402, 365–366

49 O'Brien, S.J. *et al.* (1999) The promise of comparative genomics in mammals. *Science* 286, 458–481

50 Darwin, C. (1859) *On the Origin of Species* (2nd edn), John Murray

Other trends

– articles of ecological or evolutionary interest in recent issues of other *Trends* and *Current Opinions* magazines

- Behavioral neuroscience: challenges for the era of molecular biology, *Israel Lederhendler* and *Jay Schulkin* **Trends in Neurosciences** 23, 451–454 (October 2000)
- Evolution of transcriptional regulation, *Diethard Tautz* **Current Opinion in Genetics and Development** 10, 575–579 (October 2000)
- On the conservation of protein sequences in evolution, *Brigitte Kisters-Woike, Catherine Vangierdegom* and *Benno Müller-Hill* **Trends in Biochemical Sciences** 25, 419–421 (September 2000)
- Intraspecies variation in bacterial genomes: the need for a species genome concept, *Ruiting Lan* and *Peter R. Reeves* **Trends in Microbiology** 8, 396–401 (September 2000)
- Evolution of eyes, *Russell D. Fernald* **Current Opinion in Neurobiology** 10, 444–450 (August 2000)
- Zebrafish on the move: towards a behavior–genetic analysis of vertebrate vision, *Herwig Baier* **Current Opinion in Neurobiology** 10, 451–455 (August 2000)
- The emergence of ecology from natural history, *Keith R. Benson* **Endeavour** 24, 59–62 (June 2000)
- Why Darwin was English, *Gabriel Finkelstein* **Endeavour** 24, 76–78 (June 2000)