Rare Genomic Changes

10.1 The Perfect Phylogenetic Marker – 196

10.2 Mobile Elements – 198

10.3 MicroRNAs – 201

10.4 Introns – 202

10.5 Gene Order – 203

10.6 Changes in the Genetic Code – 206

References – 207

C. Bleidorn, *Phylogenomics*, DOI 10.1007/978-3-319-54064-1_10

- Several different marker systems employing genome-level character have been used to find additional support for phylogenetic hypothesis.
- Genome-level characters include absence/presence patterns of mobile elements, microRNAs and introns, as well as gene order rearrangement and changes in the genetic code.
- Retrotransposon integrations spread by a copy-and-paste mechanism through the genome and are close to a perfect phylogenetic marker for shallow phylogenies (divergences of <65 mya).
- Absence/presence of microRNAs can be used to resolve deep phylogenies, but frequent convergent loss makes analyses difficult.
- Several mechanisms (inversion, transposition, tandem duplication random loss, translocation, fusion, fission) can result in the rearrangement of gene order.
- Maximum parsimony variants can be used to analyse absence/presence matrices to reconstruct phylogenetic trees.

10.1 The Perfect Phylogenetic Marker

The ideal phylogenetic marker is a character that, after it has evolved, will not be lost again, and homology can unambiguously be assigned across taxa due to its conservation. DNA or amino acid sequences are far from being perfect markers, and many problems can arise in phylogenetic analyses (Jeffroy et al. 2006). As an alternative, genome-level characters became popular to complement existing phylogenetic analyses and to test hypotheses with an independent set of characters (Rokas and Holland 2000). Possible phylogenetic markers are integrations of mobile elements, absence/presence (a/p) of microRNAs or introns, gene order rearrangements or changes in the genetic code. A big difference between these kinds of markers in comparison to analysing sequence data is how they are expected to change over time. For sequence data usually a clocklike change is assumed with the expectation that over time the numbers of changes accumulate linearly, even though the pace of change might be different in different lineages. In contrast, genome-level characters are expected to change non-clocklike in a saltatory way (Boore 2006). This makes analysing rare genomic change data tricky, as evolutionary models are more difficult to apply. However, if the changes are indeed rare, the presence of such characters might be an additional strong support for the monophyly of its bearers, which could be especially interesting for clades that are difficult to resolve by sequence data alone. For example, molecular systematic analyses based on a single or few genes consistently recovered a monophyletic group including crustacean and insect taxa (Pancrustacea) (Friedrich and Tautz 1995). This result was controversial, as it contradicted the former textbook knowledge which united insects with myriapods (Tracheata). A single translocation of a tRNA in the rather conserved arthropod mitochondrial genome, which was only found in analysed insects and crustaceans, gave additional strong support for the Pancrustacea hypothesis (Boore et al. 1998), which is now generally accepted. Another famous example is the analysis of the presence of some mobile elements (SINEs) in specific positions in the genome, which supported the monophyly of whales, ruminants and hippopotamuses (Shimamura et al. 1997). These promising results spurned the search for rare genomic changes to resolve difficult phylogenetic questions, but also led to the question how to analyse these markers and how to weigh their support.

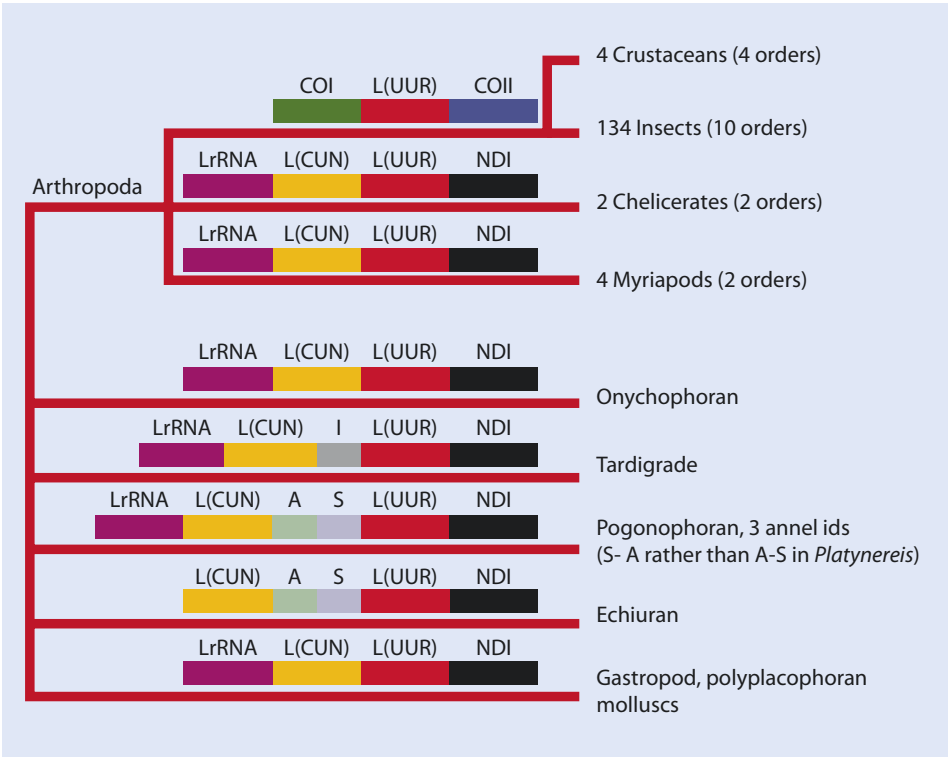


Fig. 10.1 Arthropod relationships deduced from rare genomic changes. Insects and crustaceans are united by a single tRNA (trnL) translocation within the mitochondrial genome, which is found to be (syn) apomorphic by outgroup comparison (Reprinted by permission from Macmillan Publishers Ltd.: Nature (Boore et al. 1995), copyright 1998)

Analysing genome-level characters is in the most cases based on matrices with a/p character states for the investigated taxa. This matrix can then be analysed with maximum parsimony (MP) or other approaches. As such, the analysis of genome-level characters is similar to that of morphological data. If a tree is known, characters and their states can be mapped on the phylogeny to distinguish plesiomorphic and apomorphic character states (Hennig 1965). This distinction goes back to the work of the German entomologist Willi Hennig and is seen as the foundation of the cladistic method and brought important changes of how to address phylogenetic systematics in general (Richter and Meier 1994). The plesiomorphic character state is the ancestral state present in a taxon and retained from its ancestor. For example, in Fig. 10.1 the position of the trnL(UUR) between the genes trnL(CUN) and ND1 represents a plesiomorphy, as supported by the tree and outgroup comparison. In contrast, apomorphic character states are derived states. Only these characters can be used to support the monophyly of a group of taxa. Additionally, autapomorphies (apomorphic character states found in a single lineage) and synapomorphies (apomorphic character states supporting the monophyly of a group of taxa) can be distinguished. For example, in Fig. 10.1 the position of the trnL(UUR) between the genes COI and COII is interpreted as a synapomorphy for a clade uniting insects and crustaceans. It is important to keep in mind that these terms are relative, related to where in the

phylogenetic tree they are applied. Whereas the character state for the position of trnL(UUR) is apomorphic for Pancrustacea, it is the plesiomorphic state if we would apply it to describe the same character state within the insect phylogenetic tree. In most cases plesiomorphy or apomorphy can only be assigned after (a posteriori) the phylogenetic analysis, as the characters (and the direction of their evolution) are polarized by using the resulting phylogeny. However, the presence state of some genome-level characters (e.g. SINE insertions, see below) is quasi used a priori as the apomorphic character state, as absence is unlikely the derived state (Shedlock and Okada 2000).

10.2 Mobile Elements

Retrotransposons are mobile elements that have the ability to integrate into the genome at a new site within their cell of origin (Kazazian 2004). In contrast to DNA transposons, which use a cut-and-paste mechanism of copying, retrotransposons use a copy-and-paste mechanism to integrate at new sites. This is achieved by transcription of the retrotransposons into RNA, which are then reverse transcribed and reintegrated into the genome, thereby duplicating the element. Two major classes of retrotransposons are recognized: LTR retrotransposons, which either contain long terminal repeats (LTR) at both ends, or non-LTR retrotransposons (or LINEs) which lack LTRs and possess a polyadenylate sequence at their 3'termini (Kazazian 2004). Unlike LTR retrotransposons that generate uniform target site duplications and require the presence of their terminal repeats for integration, non-LTR element copies are often truncated at their 5'ends (Malik and Eickbush 1998). Short interspersed nuclear elements (SINEs) are mobile elements that originated from the accidental retrotransposition of small RNA polymerase III transcripts, such as 7SL RNAs, tRNAs or 5S RNA. Therefore SINEs always feature an internal RNA polymerase III promoter at their 5′ end for their transcription. SINEs are nonautonomous, and to be replicated, they completely rely on the machinery of the cell and the activity of autonomous retrotransposons, such as LINEs (Kramerov and Vassetzky 2005). Some SINEs are known to occur in huge copy numbers in their «host» genome, as, for example, the primate specific *Alu* SINE family, which makes up around 11% of the human genome (Deininger 2011).

The copy-and-paste mechanism makes retrotransposons an almost perfect phylogenetic marker. They are regarded as being nearly homoplasy-free, as convergent integrations at the exact same genomic positions are highly unlikely (Ray et al. 2006), even though some few examples of apparent homoplasy exists (Han et al. 2011). And they are basically polarized characters, such that the absence of a retrotransposon at a given locus is usually the ancestral state (Ray et al. 2006). The caveat is that these markers are only suited to resolve relatively young divergences (50 mya and younger), as otherwise the homology between integrated sequences is difficult to detect as mutations are accumulated over time (Shedlock and Okada 2000). Retrotransposons have been successfully used to address population diversity in plants (Kalendar et al. 2011) or the phylogeny of birds (Suh et al. 2011) or mammals (Kriegs et al. 2006), but there are no examples to use them to infer deeper phylogenies.

Retrotransposon activity in the genome of their host varies over evolutionary timescales. Different groups of retrotransposons may have different (and also overlapping) times of activities before they get inactive. Activity of retrotransposons can be triggered by mutations within inactive sequences (e.g. acquisition of a new promoter), but also due

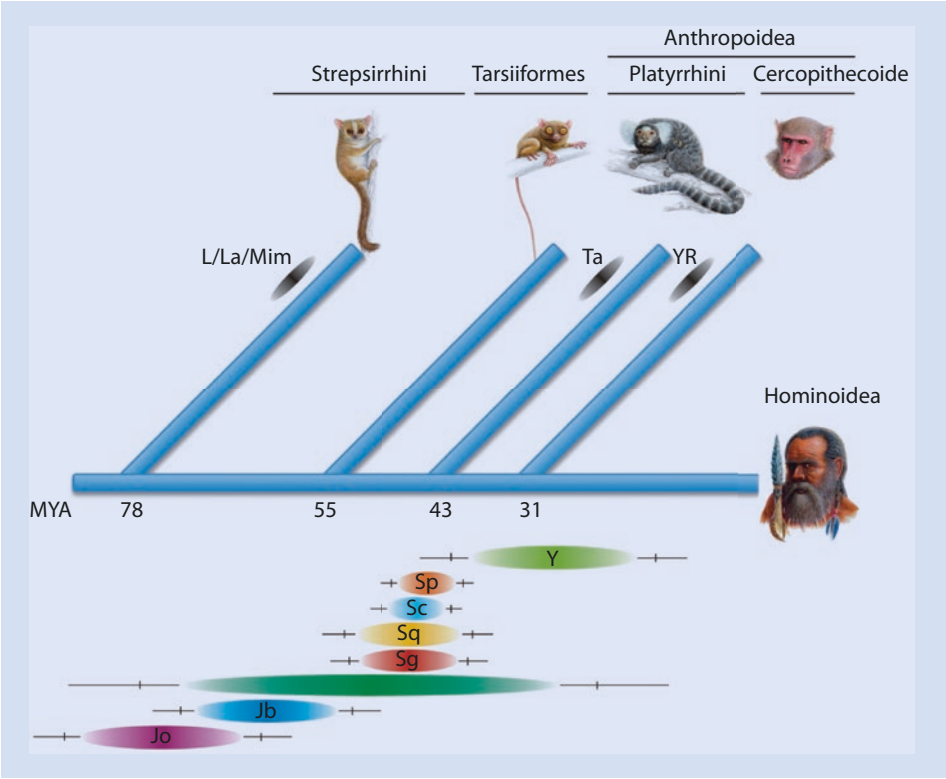


Fig. 10.2 Activity patterns of different groups of SINEs (*Jo, Jb, Sx, Sg, Sc, Sp, Y*) along primate evolution as modelled by the TinT method (Reprinted from Churakov et al. (2010))

to horizontal gene transfer into a new host (Huang et al. 2012). Inactivity or death of a group of retrotransposons occurs when the last active copy loses its activity due to mutation. Groups of retrotransposons can be classified by its sequence similarity, e.g. using the software REPEATMASKER (Tarailo-Graovac and Chen 2009), and some methods exist to trace their activity over time (Kriegs et al. 2007; Giordano et al. 2007). The «Transposition in Transposition» (TinT) method is based on the idea that evolutionary younger actively transposed elements are able to insert into older elements, but the opposite is not possible. Within a probabilistic framework, information of the occurrence of nested retrotransposon insertion patterns is used to model the timing of element activity (Churakov et al. 2010). Absolute timescales of the relative chronological order can be inferred by mapping these activity patterns on a dated phylogeny (Fig. 10.2), which also highlights that retrotransposons are often only informative for a short window of the evolutionary timescale. For example, analyzing *Jb* SINE patterns would not be informative to investigate ape (Hominoidea) evolution, as its activity window predates the origin of this clade (Fig. 10.2).

For phylogenetic analyses, a/p of retrotransposons is scored for each homologous integration within a character matrix. Especially integrations of retrotransposons within intron regions are suitable for analysis, as due to the conserved nature of the adjacent exons these genomic regions are easier to orthologize. Besides the orthology of the genomic region, homology of the retrotransposons must be carefully considered, which is

complicated due to random mutational decay over time. It is not unusual to manually inspect every single alignment of orthologous genes to verify the homology of retrotransposons (Suh et al. 2015). There are two strategies for the phylogenetic analysis of retrotransposon data. First, a/p matrices can be analysed directly using MP for the inference of a tree, e.g. Kaiser et al. (2007). However, reflecting their activity sometimes these markers are only informative for certain windows of evolutionary time (see above), and several parts of the tree remain unresolved. Alternatively, presence of shared retrotransposon integrations is mapped onto a tree topology and congruence with a/p patterns can be used to favour one of several competing hypotheses or to give additional support for the monophyly of selected groups in a tree. For example, Kriegs et al. (2006) analysed retrotransposon integrations across mammals and mapped their data on existing trees. A statistic framework for evaluating support from retrotransposons has been proposed by Waddell et al. (2001). Their likelihood-based test statistics show that at least five unambiguous markers (five retrotransposon integrations supporting a given clade, with no other integration in conflict) are required for a certain node to gain significant p-values.

Even though retrotransposon markers usually show only very low levels of homoplasy from convergent integration at the exactly same site, several examples of conflicting nodes have been found when addressing the phylogeny of fast radiations. For example, Nishihara et al. (2009) investigated retrotransposon integrations of placental mammals and found nearly the same number of loci (21–25 loci) supporting three different hypotheses. Similarly, Suh et al. (2015) investigated the radiation of birds based on thousands of carefully selected retrotransposons and found that a third of these are supporting conflicting hypotheses. In both analyses the conflicting retrotransposons map to parts of the phylogeny which are characterized by short internodes. Consequently, the conflict within this retrotransposon data is not interpreted as convergence due to parallel integrations, but as a persistence of ancestral polymorphisms, a phenomenon known as incomplete lineage sorting (ILS). The affected regions of the tree have been found as notoriously difficult to reconstruct, even in the light of massive datasets. High amounts of ILS in combination with short internodes point to a nearly simultaneous divergence of deep lineages within mammals and birds, which might be unresolvable into a bifurcating tree. Instead, a phylogenetic network illustrating the conflict within this part of the tree seems to be a better representation of the phylogenetic relationships of these groups (Suh et al. 2015; Hallström and Janke 2010). Not surprisingly, high amounts of ILS based on retrotransposon data has been also reported for Lake Tanganyika cichlids (Takahashi et al. 2001), the posterchild for adaptive radiations.

Mobile elements in general have been firstly described in plants (McClintock 1950), a discovery which later was honoured with the Nobel Prize in physiology or medicine in 1983 for Barbara McClintock. And even though different classes of mobile elements are extremely abundant in plant genomes, most studies exploiting these elements as phylogenetic markers are from vertebrate animals. Kalendar et al. (2011) summarized the use of mobile elements as markers in plant phylogeny and evolution, with most studies addressing the population level. Yaakov et al. (2012) used so-called miniature inverted-repeat transposable elements (MITEs) to investigate wheat biodiversity and evolution. MITEs are small mobile elements of up to a few hundred base pairs in size, flanked by tandemly inverted repeats (Wicker et al. 2007). Similarly to SINEs, they are nonautonomous and can occur in high copy numbers. MITEs were first discovered in plants (Wessler et al. 1995), but are also abundantly found in many eukaryotic genomes, including humans (Morgan 1995). The study by Yaakov et al. (2012) analysed a/p matrices of MITE-polymorphism

across wheat accessions and found this marker system phylogenetically informative. Similar positive results were reported for the analyses of barley biodiversity using MITEs (Lyons et al. 2008).

10.3 MicroRNAs

MicroRNAs are short non-coding RNAs involved in the regulation of gene expression. They are found in plants and animals, but seem to have evolved convergently in these lineages (Shabalina and Koonin 2008). Moreover, several viruses have been identified harbouring microRNA sequences, with most examples stemming from herpesviruses (Skalsky and Cullen 2010). Premature microRNAs form hairpin-like secondary structures (Fig. 10.3). This stem-loop precursor is further processed into double-stranded RNA of approximately 22 bp (Kim 2005). Nucleotides 2–7 of the mature microRNA 5′-end are called seeds and play an important role in mRNA-target recognition (Liu et al. 2008). Often more than a hundred targets can be recognized by a single microRNA, and miRNAs complementarily bind to target mRNAs, where they repress translation and/or induce mRNA degradation (Lewis et al. 2003). Mature microRNAs were shown to be highly conserved across animal taxa (Sempere et al. 2006), and several hundred distinct microRNA families have been reported for Metazoa (Kozomara and Griffiths-Jones 2011).

Interestingly, it has been found that microRNA families are continuously emerging and expanding through animal evolution (Hertel et al. 2006); however, once evolved microRNAs were regarded to be rarely lost secondarily (Sempere et al. 2006). The expression of many microRNAs is known to be tissue specific (Clark et al. 2010; Christodoulou et al. 2010), and additionally, the disparity of microRNAs of a given animal taxon can often be linked to its morphological complexity (e.g. number of cell types) (Heimberg et al. 2008; Sempere et al. 2006; Kosik 2009). Given these properties, the potential of microRNAs as a powerful phylogenetic marker system is obvious and they were used in several phylogenetic analyses (Heimberg et al. 2010; Sperling et al. 2011; Rota-Stabelli et al. 2011; Sperling et al. 2009; Campbell et al. 2011; Helm et al. 2012).

As in the case of retrotransposons, microRNAs are coded as a/p in a matrix and can be analysed directly (e.g. using MP) or by mapping onto an existing phylogenetic tree. The advantage of microRNAs over retrotransposons is that they have shown to be

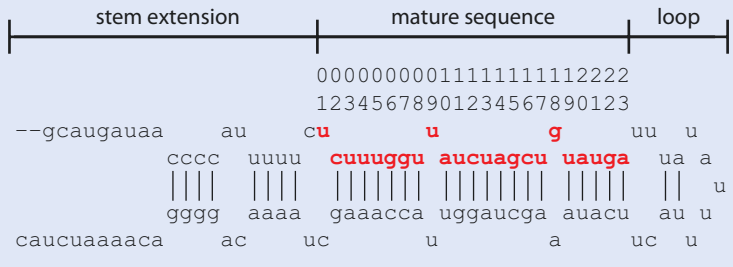


Fig. 10.3 Features of premature microRNA secondary structure exemplified by *mir-9* from the annelid *Capitella teleta* as found in miRBase (Kozomara and Griffiths-Jones 2014) accession MI0010052. The mature sequence is indicated by red bases. Positions 2–8 of the mature sequence are also known as seed and play an important role in mRNA-target recognition

phylogenetically informative over deep evolutionary timescales (Tarver et al. 2013). Even though mature microRNAs are represented by very small sequences (~22 bp), they remain remarkably conserved over time. However, in contrast to analyses of retrotransposon data, phylogenetic analyses of microRNAs seem to be more problematic. Based on the presence of a single or few microRNAs, Acoela are supported as a deuterostome in-group taxon (Philippe et al. 2011), Mandibulata as monophyletic (Rota-Stabelli et al. 2011) and Tardigrada as sister group of Onychophora + Arthropoda (Campbell et al. 2011). However, these results might be biased due to highly heterogeneous rates of microRNA gain and loss, as well as sampling error (Thomson et al. 2014). For example, there is evidence that microRNAs get lost due to loss of their function. For example, *mir-10* is a phylogenetically conserved microRNA present in most bilaterian lineages regulating a subset of *hox* genes (Pearson et al. 2005). Interestingly, this microRNA is directly located within the *hox* cluster and has been convergently lost in lineages with a disintegrated *hox* cluster, such as nematodes and tunicates (Tanzer et al. 2005). Major loss of microRNAs is also reported for tunicates (Fu et al. 2008), where at least 11 families of bilaterian microRNAs are missing. An analysis of chordate phylogeny also revealed several losses of microRNA families in different lineages (Heimberg et al. 2010). Frequent gain and loss on a short evolutionary timescale has further been demonstrated for *Drosophila* (Nozawa et al. 2010). Moreover, it remains a practical problem that the absence of microRNAs can be only safely concluded with available complete genome sequences. Instead of MP analyses or mapping, Thomson et al. (2014) explored the performance of microRNA a/p matrices under different evolutionary models. Their re-analyses casted doubt on the results of several published phylogenetic studies, and they conclude that the potential of microRNA data to resolve the (animal) tree of life has been overstated.

10.4 Introns

In eukaryotes genes are interrupted by spliceosomal introns which are removed from transcripts prior to their translation (Jeffares et al. 2006). The absolute number of introns within a genome and the number of introns within a gene are highly variable. However, intron positions of most introns are conserved across eukaryotes, and variation in intron numbers is explained by either intron gain or loss (Rogozin et al. 2003). Possible sources for the generation of new genomic introns are DNA transposons (Huff et al. 2016). Generally, intron gain seems to occur less frequently; however, increased intron gain and decreased intron loss was observed in evolutionarily conserved genes (Carmel et al. 2007). Moreover, intron loss is regarded to be nearly irreversible (Roy and Gilbert 2005a), as intron gain at exactly the same site happens only rarely (Sverdlov et al. 2005). Therefore, shared intron positions should indicate homology (Roy 2016), and analysing a/p patterns of intron positions has been proposed as a useful phylogenetic marker of deep divergences (Rokas and Holland 2000). For example, introns have been used to analyse relationships of deep divergences within Metazoa (Roy and Gilbert 2005b) or ray-finned fishes (Venkatesh et al. 1999).

A straightforward way to analyse a/p data of introns is based on MP. In its general form, MP gives all character transformations the same probability. However, as intron loss is thought to be nearly irreversible, this assumption might be violated. To circumvent this problem, the use of a special form of MP called Dollo parsimony can be used to analyse such datasets and is also applicable for microRNA data. Dollo's Law states that complex

characters cannot be «re-evolved» once they got lost, and instead alternative ways lead to convergent solutions (which should be detectable as such) (Dollo 1893). This century-old idea has to be treated carefully given the actual knowledge of the genetic and developmental bases of complex morphological characters (Hall 2003), but might be a fitting description for what we know about the evolution of microRNAs and introns. Dollo parsimony was introduced by Farris (1977). For analysis, characters are polarized a priori, and the presence of the complex character state is coded as 1, whereas the absent, likely ancestral state is coded by 0. Using this algorithm, only one change from 0 to 1 is allowed during the analysis, whereas as many reversions from 1 to 0 as necessary to explain the observed data are possible. By applying Dollo parsimony for intron a/p matrices, the number of (parallel) intron gains is minimized, whereas losses can be frequent. As the rates of intron loss can vary dramatically across taxa (Jeffares et al. 2006), Zheng et al. (2007) introduced a modified Dollo parsimony algorithm that uses different weights for the cost of an intron loss in different branches. Alternatively, explicit phylogenetic models have been developed to analyse large matrices of intron a/p data across species. In this case, based on different tree topologies (hypotheses), different expectations regarding ratios of intron gain and lost are formulated and compared with the data (Roy and Gilbert 2005b). Both types of analyses implicitly assume constant rates of intron loss and violations of this assumption might lead to long-branch attraction (Irimia and Roy 2008). Especially the frequent occurrence of multiple independent losses of the same intron in distantly related species is problematic for phylogenetic analyses and questions the usefulness of this marker system in general (Krzywinski and Besansky 2002; Kiontke et al. 2004).

An approach to limit the impact of convergent intron gains or losses is the analysis of near intron pairs (NIPs) (Krauss et al. 2008). Such NIPs include two intron positions in an alignment of orthologous genes that are separated by a small number of nucleotides. It is known that exons smaller than ~50 bp are only rarely found, which could be related to problems of splicing such small sequences (Irimia and Roy 2008). Therefore, introns found at nearby positions are unlikely to have coexisted, and given that multiple gains at exactly the same site are very rare, this data can be used to infer a phylogenetic tree. Krauss et al. (2008) proofed that this method is in principle useful for phylogenetic reconstruction. Lehmann et al. (2013) used NIPs to infer metazoan phylogeny and found them to clearly outperform Dollo parsimony analyses based on all introns. However, as the number of suitable characters is strongly reduced by this approach, parts of the tree which correspond to taxa or time periods with low levels of intron gain are difficult to resolve.

10.5 Gene Order

The order of genes in the genome has been extensively used as a phylogenetic marker (Boore 2006; Sankoff et al. 1992). The first use of gene order to infer evolutionary relationships goes back to Sturtevant and Dobzhansky (1936), who analysed inversions located in a chromosome to study the evolution of some drosophilids. Most studies using gene order as phylogenetic markers are based on organellar genomes, as in the case of plant chloroplasts (Downie and Palmer 1992; Cosner et al. 2004) or animal mitochondrial genomes (Boore and Brown 1998; Bleidorn et al. 2007). For example, animal mitochondrial genomes are usually circular molecules that harbour around 37 genes. Every gene can be either transcribed from the plus or the minus strand, and several mechanisms have been described how gene order can be rearranged (Boore 1999). Due to its small size, many

animal mitochondrial genomes have been sequenced already with the Sanger technique. Using next-generation sequencing techniques, complete animal mitochondria can now be reconstructed fast and easily from shallowly sequenced whole genome shotgun libraries, an approach which is known as genome skimming (Richter et al. 2015). Similarly, complete chloroplast genome have been reconstructed using this approach (Malé et al. 2014). Not surprisingly, many different mitochondrial gene orders are observed (e.g. Fig. 10.4), and the possibility of convergent changes resulting in the same order is rather low (Dowton et al. 2002), even though some examples are known (Shao and Barker 2003).

Several types of rearrangements are defined based on the comparison of closely related species with different gene orders of unichromosomal genomes: inversions (Fig. 10.5a), transpositions (Fig. 10.5b), inverse transpositions (transpositions where the re-inserted fragment is inverted) (Fig. 10.5c) and tandem duplications followed by random loss (TDRL) of one of the gene copies (Fig. 10.5d) (Bernt et al. 2013). A web-based application called CREX (Bernt et al. 2007) is available, which based on common intervals finds parsimonious scenarios for the rearrangement of a pair of gene orders. More complicated are cases where more than one chromosome exists, as, for example, in most eukaryotic

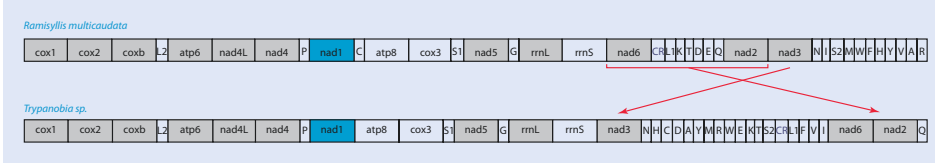


Fig. 10.4 Comparison of the mitochondrial gene order of two closely related annelids (Syllidae). All genes are transcribed from the same strand. Changes are indicated by *arrows*. Mitochondrial protein coding and ribosomal genes are abbreviated with 3–4 letters, tRNA genes are given in the one-letter code (Reprinted from Aguado et al. (2015))

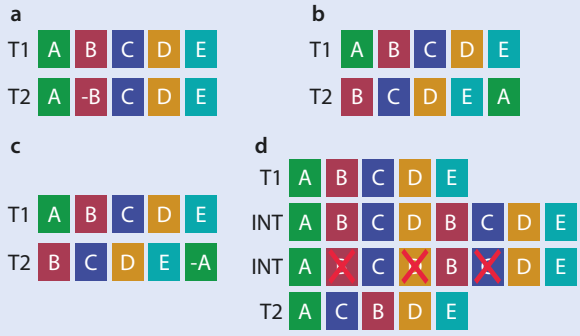
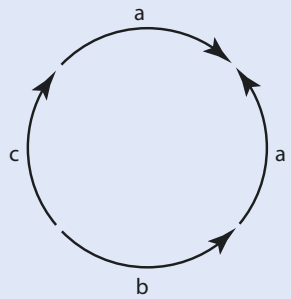


Fig. 10.5 Different types of rearrangements hypothesized for animal mitochondrial genomes. Examples of gene order of five genes (A-E) in two taxa (T1 and T2). Genes on the minus strand are indicated with a minus sign (–). **a** Inversion of a single gene (B to –B). **b** Transposition of a single gene (A). **c** Inverse transposition of a single gene (A to –A). **d** Tandem duplication random loss (TDRL) scenario. Change of the gene order from T1 to T2 is shown with intermediate states (INT), where three genes (B, C, D) are tandemly duplicated and one copy of each gene got lost. Sometimes different scenarios (e.g. transposition and TDRL) are equally likely to explain changes in the gene order

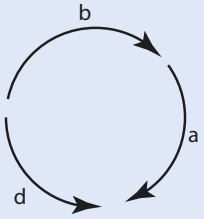
nuclear genomes. In this case, translocations, fusions and fissions are additional possible scenarios (Hu et al. 2014). A translocation describes the break of a chromosome, with one part attaching to another chromosome. A fusion joins two chromosomes, while a fission breaks a single chromosome into two parts. Analysing gene orders is further complicated by the deletion and duplication of genes, and in the latter case even whole genome duplications are not unusual.

Different methods have been proposed to analyse gene order data. Genes can be coded on different strands of a double-stranded DNA molecule, thereby allowing four different types of adjacency (Moret et al. 2013) of two genes following each other: (a^t, b^t), (a^h, b^t), (a^t, b^h) and (a^h, b^h). In this examples, *a* and *b* denote different genes, whereas ^{h} (head) and ^{t} (tail) refer to their orientation to each other. For example, when the two genes *b* and *a* are described on the plus strand one after another, their orientation would be (b^h, a^t) (Fig. 10.6a, genome 2). The easiest way to compare the order of two genomes is to estimate the number of breakpoints (Blanchette et al. 1997). If two genes *a* and *b* are adjacent in taxon 1 but not in taxon 2, they determine a breakpoint. The number of breakpoints between two unichromosomal genomes represents the most general measure of gene order distance, as it requires no assumptions about the mechanisms of gene order evolution (e.g. differences between inversions and transpositions). Further distance measures are the inversion distance and the double-cut-and-join (DCJ) distance. The inversion distance equals the minimum number of inversions to transform one unichromosomal gene order into another one, given the same gene content and absence of duplications (Hannenhalli and Pevzner 1999). The (DCJ) distance is a model that accounts for most events altering gene order, such as inversions, translocations, fusions and fissions

a



Genome 1



Genome 2

b

	adjacency information						content information			
	$\{a^h, a^h\}$	$\{a^t, b^h\}$	$\{a^t, c^h\}$	$\{b^t, c^t\}$	$\{a^h, d^h\}$	$\{b^t, d^t\}$	a	b	c	d
Genome 1	1	1	1	1	0	0	1	1	1	0
Genome 2	0	1	0	0	1	1	1	1	0	1

Fig. 10.6 Two circular example genomes coded for gene order analysis. **a** Genome 1 and genome 2, showing the order of genes (a–d). *Arrows* show the orientation of the gene. **b** Coding of the two genomes into a matrix with adjacency and content information. The orientation of genes is given as head to head (e.g. a^h, a^h), or tail to head (e.g. a^t, b^h), and so on. Presence of an adjacency is coded as 1, absence as 0. Similarly, gene content information is coded as absent (1) or present (0) (Reprinted from Lin et al. (2012a))

(Yancopoulos et al. 2005). Calculations of different pairwise distances between genomes can be conducted with the software UNIMOG (Hilker et al. 2012). Whereas most distance measures were developed for unichromosomal genomes with the exact same gene content, alternatives are available for mutichromosomal genomes and/or when gene duplications or deletions occurred (Moret et al. 2013). Instead of using distance measures, gene order can be also coded into a matrix, where each observed combination of adjacent genes (and their orientation) represents a character, and the absence or presence of the adjacency of these genes is coded for each genome (Fig. 10.6). Phylogenetic analyses of gene order data can be either conducted using distance-based methods such as neighbour joining or by analysing encoded gene order matrices using an optimality criterion such as MP or maximum likelihood (ML) (Moret et al. 2013). Several programs specifically for the phylogenetic analysis of gene order data have been published, e.g. GRAPPA (Moret et al. 2001), MGR (Bourque and Pevzner 2002), MLGO (Hu et al. 2014) or TIBA (Lin et al. 2012b). Lin et al. (2012a) developed a likelihood approach where gene order and content are coded into a matrix (see Fig. 10.6) and transition probabilities between character states are estimated from this matrix, which can then be used for ML analyses, e.g. by using the program RAXML (Stamatakis 2014). Matrices based on pairwise distances derived from gene order can be further analysed using neighbour joining, e.g. as implemented in MEGA (Kumar et al. 2016).

10.6 Changes in the Genetic Code

After the structure of the DNA double helix was discovered in 1953, it took more than a decade to completely decipher its code (Cobb 2015). Based on this code, nucleotide triplets are translated into amino acids. When discovered, it was surprising that the code was highly degenerated, as most of the 20 amino acids were represented by more than one triplet. Initially, it was considered that the genetic code is truly universal and not evolvable, meaning that the pattern of degeneracy could represent a «frozen accident» (Crick 1968). However, after the discovery that human nuclear and mitochondrial genes use different codes (Barrell et al. 1979), it became obvious that the code is indeed evolvable. Later on, many exceptions from the standard genetic code have been described, with most of them found in mitochondrial genomes (Knight et al. 2001). Different models have been proposed how codons can be reassigned. Based on the codon-capture model it is hypothesized that a codon first disappears from the coding sequences of the genome, resulting into loss of function of this specific codon. In case it reappears due to nucleotide substitutions in any coding sequence, a reassignment to a new tRNA is possible (Osawa and Jukes 1989). Alternatively, a codon might be translated ambiguously, and one of its variant becomes fixed (Schultz and Yarus 1994). There are examples available for both models, which may just represent differences in the timing if reassigned codons appear after or before the loss of the old codon (Sengupta et al. 2007).

Given that changes in the genetic code are rare events, they bear the potential to be used as a phylogenetic marker. By comparatively analysing mitochondrial genomes, Castresana et al. (1998) found support for the monophyly of a group uniting enteropneusts and echinoderms based on predicted changes of the genetic code. Similarly, Telford et al. (2000) used a change in the genetic code as further support for the monophyly of the flatworm taxon Rhabditophora, whereas Keeling and Doolittle (1997) used such data to

evaluate different hypotheses regarding the phylogenetic position of the taxon *Girardia* within diplomonads.

As the loss of a codon triplet sequence in the coding part of the genome is an important step towards codon reassignment, it comes without surprise that most code changes have been reported from the rather small animal mitochondrial genomes. Abascal et al. (2012) screened more than 300 arthropod mitochondrial genomes and found that ~20% do not bear the codon AGG. Interestingly, in nearly half of the investigated species, this codon is translated into lysine, whereas the other half shows a translation into leucine. When mapping these changes onto an arthropod phylogeny, it became clear that a reassignment of this codon occurred frequently within this group, exhibiting high levels of convergence, thereby diminishing its usefulness as a phylogenetic character. The same authors also published a software called GENDECODER which can be used to automatically scan genomes for the presence of reassigned codons (Abascal et al. 2006). This method is based on the idea that if the appearance of a particular codon in an investigated species is linked to an alignment position for which a specific amino acid is conserved in a set of reference species, the same translation is assumed for the query. Alternatively, the software FACIL uses hidden Markov models to predict genetic codes by comparison with a reference database (Dutilh et al. 2011).

References

Abascal F, Posada D, Zardoya R (2012) The evolution of the mitochondrial genetic code in arthropods revisited. Mitochondr DNA 23:84–91

Abascal F, Zardoya R, Posada D (2006) GenDecoder: genetic code prediction for metazoan mitochondria. Nucleic Acids Res 34:W389–W393

Aguado MT, Glasby CJ, Schroeder PC, Weigert A, Bleidorn C (2015) The making of a branching annelid: an analysis of complete mitochondrial genome and ribosomal data of *Ramisyllis multicaudata*. Sci Rep 5:12072

Barrell BG, Bankier AT, Drouin J (1979) A different genetic code in human mitochondria. Nature 282:189–194

Bernt M, Braband A, Schierwater B, Stadler PF (2013) Genetic aspects of mitochondrial genome evolution. Mol Phylogenet Evol 69:328–338

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M (2007) CREx: inferring genomic rearrangements based on common intervals. Bioinformatics 23:2957–2958

Blanchette M, Bourque G, Sankoff D (1997) Breakpoint phylogenies. Genome Inform Ser Workshop Genome Inform 8:25–34

Bleidorn C, Eeckhaut I, Podsiadlowski L, Schult N, McHugh D, Halanych KM, Milinkovitch MC, Tiedemann R (2007) Mitochondrial genome and nuclear sequence data support Myzostomida as part of the annelid radiation. Mol Biol Evol 24:1690–1701

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780

Boore JL (2006) The use of genome-level characters for phylogenetic reconstruction. Trends Ecol Evol 21:439–446

Boore JL, Brown WM (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev 8:668–674

Boore JL, Collins T, Stanton D, Daehler L, Brown WM (1995) Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature 376:163–165

Boore JL, Lavrov DV, Brown WM (1998) Gene translocation links insects and crustaceans. Nature 392:667–668

Bourque G, Pevzner PA (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Res 12:26–36

Campbell LI, Rota-Stabelli O, Edgecombe GD, Marchioro T, Longhorn SJ, Telford MJ, Philippe H, Rebecchi L, Peterson KJ, Pisani D (2011) MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. Proc Natl Acad Sci U S A 108:15920–15924

Carmel L, Rogozin IB, Wolf YI, Koonin EV (2007) Evolutionarily conserved genes preferentially accumulate introns. Genome Res 17:1045–1050

Castresana J, Feldmaier-Fuchs G, S-i Y, Satoh N, Pääbo S (1998) The mitochondrial genome of the hemichordate *Balanoglossus carnosus* and the evolution of deuterostome mitochondria. Genetics 150:1115–1123

Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D (2010) Ancient animal microRNAs and the evolution of tissue identity. Nature 463:1084–1088

Churakov G, Grundmann N, Kuritzin A, Brosius J, Makałowski W, Schmitz J (2010) A novel web-based TinT application and the chronology of the Primate Alu retroposon activity. BMC Evol Biol 10:376

Clark AM, Goldstein LD, Tevlin M, Tavare S, Shaham S, Miska EA (2010) The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans*. Nucleic Acids Res 38:3780–3793

Cobb M (2015) Life's greatest secret: the race to crack the genetic code. Basic Books, London

Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol Biol 4:27

Crick FHC (1968) The origin of the genetic code. J Mol Biol 38:367–379

Deininger P (2011) Alu elements: know the SINEs. Genome Biol 12:236

Dollo L (1893) Les lois de l' evolution. Bull Belg Soc Geol, Palaeontol Hydrol 8:164–166

Downie SR, Palmer JD (1992) Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Springer US, Boston, pp 14–35

Dowton M, Castro LR, Austin AD (2002) Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome 'morphology'. Invertebr Syst 16:345–356

Dutilh BE, Jurgelenaite R, Szklarczyk R, van Hijum SAFT, Harhangi HR, Schmid M, de Wild B, Françoijs KJ, Stunnenberg HG, Strous M, Jetten MSM, Op den Camp HJM, Huynen MA (2011) FACIL: fast and accurate genetic code inference and logo. Bioinformatics 27:1929–1933

Farris JS (1977) Phylogenetic analysis under Dollo's law. Syst Zool 26:77–88

Friedrich M, Tautz D (1995) Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature 376:165–167

Fu XH, Adamski M, Thompson EM (2008) Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. Mol Biol Evol 25:1067–1080

Giordano J, Ge Y, Gelfand Y, Abrusán G, Benson G, Warburton PE (2007) Evolutionary history of mammalian transposons determined by genome-wide defragmentation. PLoS Comput Biol 3:e137

Hall BK (2003) Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. Biol Rev 78:409–433

Hallström BM, Janke A (2010) Mammalian evolution may not be strictly bifurcating. Mol Biol Evol 27:2804–2816

Han K-L, Braun EL, Kimball RT, Reddy S, Bowie RCK, Braun MJ, Chojnowski JL, Hackett SJ, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T (2011) Are transposable element insertions homoplasy free?: An examination using the Avian tree of life. Syst Biol 60:375–386

Hannenhalli S, Pevzner PA (1999) Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals. J ACM 46:1–27

Heimberg AM, Cowper-Sallari R, Semon M, Donoghue PCJ, Peterson KJ (2010) microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. Proc Natl Acad Sci U S A 107:19379–19383

Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ (2008) MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci U S A 105:2946–2950

Helm C, Bernhart SH, Siederdissen CHZ, Nickel B, Bleidorn C (2012) Deep sequencing of small RNAs confirms an annelid affinity of Myzostomida. Mol Phylogenet Evol 64:198–203

Hennig W (1965) Phylogenetic systematics. Annu Rev Entomol 10:97–116

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF, Students Bioinformatics Computer L (2006) The expansion of the metazoan microRNA repertoire. BMC Genomics 7:25

Hilker R, Sickinger C, Pedersen CNS, Stoye J (2012) UniMoG—a unifying framework for genomic distance calculation and sorting based on DCJ. Bioinformatics 28:2509–2511
Hu F, Lin Y, Tang J (2014) MLGO: phylogeny reconstruction and ancestral inference from gene-order data. BMC Bioinformatics 15:354
Huang CRL, Burns KH, Boeke JD (2012) Active transposition in genomes. Annu Rev Genet 46:651–675
Huff JT, Zilberman D, Roy SW (2016) Mechanism for DNA transposons to generate introns on genomic scales. Nature 538:533–536
Irimia M, Roy SW (2008) Spliceosomal introns as tools for genomic and evolutionary analysis. Nucleic Acids Res 36:1703–1712
Jeffares DC, Mourier T, Penny D (2006) The biology of intron gain and loss. Trends Genet 22:16–22
Jeffroy O, Brinkmann H, Delsuc F, Philippe H (2006) Phylogenomics: the beginning of incongruence? Trends Genet 22:225–231
Kaiser VB, van Tuinen M, Ellegren H (2007) Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in Galliform Birds. Mol Biol Evol 24:338–347
Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C, Schulman AH (2011) Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106:520–530
Kazazian HH (2004) Mobile elements: drivers of genome evolution. Science 303:1626–1632
Keeling PJ, Doolittle WF (1997) Widespread and ancient distribution of a noncanonical genetic code in diplomonads. Mol Biol Evol 14:895–901
Kim VN (2005) MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol 6:376–385
Kiontke K, Gavin NP, Raynes Y, Roehrig C, Piano F, Fitch DHA (2004) *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc Natl Acad Sci U S A 101: 9003–9008
Knight RD, Freeland SJ, Landweber LF (2001) Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2:49–58
Kosik KS (2009) OPINION MicroRNAs tell an evo-devo story. Nat Rev Neurosci 10:754–759
Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39:D152–D157
Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73
Kramerov DA, Vassetzky NS (2005) Short retroposons in Eukaryotic genomes. In: International review of cytology, vol 247. Academic Press, New York, pp 165–221
Krauss V, Thümmler C, Georgi F, Lehmann J, Stadler PF, Eisenhardt C (2008) Near intron positions are reliable phylogenetic markers: an application to holometabolous insects. Mol Biol Evol 25:821–830
Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J (2006) Retroposed elements as archives for the evolutionary history of placental mammals. PLoS Biol 4:e91
Kriegs JO, Matzke A, Churakov G, Kuritzin A, Mayr G, Brosius J, Schmitz J (2007) Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). BMC Evol Biol 7:190
Krzywinski J, Besansky NJ (2002) Frequent intron loss in the white gene: a cautionary tale for phylogeneticists. Mol Biol Evol 19:362–366
Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
Lehmann J, Stadler PF, Krauss V (2013) Near intron pairs and the metazoan tree. Mol Phylogenet Evol 66:811–823
Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115:787–798
Lin Y, Fei H, Tang J, Moret BME 2012a Maximum likelihood phylogenetic reconstruction from high-resolution whole-genome data and a tree of 68 eukaryotes. In: Proceedings of the 18th pacific symposium on Biocomputing (PSB'13), Singapore, World Scientific, pp 285–296
Lin Y, Rajan V, Moret BME (2012b) TIBA: a tool for phylogeny inference from rearrangement data with bootstrap analysis. Bioinformatics 28:3324–3325
Liu N, Okamura K, Tyler DM, Phillips MD, Chung WJ, Lai EC (2008) The evolution and functional diversification of animal microRNA genes. Cell Res 18:985–996
Lyons M, Cardle L, Rostoks N, Waugh R, Flavell AJ (2008) Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome. Mol Gen Genomics 280:275–285

Malé P-JG, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, Lhuillier E, Scotti-Saintagne C, Tinaut A, Chave J (2014) Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol Ecol Resour 14:966–975

Malik HS, Eickbush TH (1998) The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol Biol Evol 15:1123–1134

McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A 36:344–355

Moret BME, Lin Y, Tang J (2013) Rearrangements in phylogenetic inference: compare, model, or encode? In: Chauve C, El-Mabrouk N, Tannier E (eds) Models and algorithms for genome evolution. Springer London, London, pp 147–171

Moret BME, Wyman S, Bader D, Warnow T, Yan M A 2001 new implementation and detailed study of breakpoint analysis. In: Proceedings of the 6th pacific symposim on Biocomputing (PSB'01), Singapore, World Scientific, pp 583–594

Morgan GT (1995) Identification in the human genome of mobile elements spread by DNA-mediated transposition. J Mol Biol 254:1–5

Nishihara H, Maruyama S, Okada N (2009) Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. Proc Natl Acad Sci U S A 106:5235–5240

Nozawa M, Miura S, Nei M (2010) Origins and evolution of microRNA genes in *Drosophila* species. Genome Biol Evol 2:180–189

Osawa S, Jukes TH (1989) Codon reassignment (codon capture) in evolution. J Mol Evol 28:271–278

Pearson JC, Lemons D, McGinnis W (2005) Modulating Hox gene functions during animal body patterning. Nat Rev Genet 6:893–904

Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, Wallberg A, Peterson KJ, Telford MJ (2011) Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. Nature 470:255–258

Ray DA, Xing J, Salem AH, Batzer MA (2006) SINEs of a nearly perfect character. Syst Biol 55:928–935

Richter S, Meier R (1994) The development of phylogenetic concepts in Hennig's early theoretical publications (1947–1966). Syst Biol 43:212–221

Richter S, Schwarz F, Hering L, Böggemann M, Bleidorn C (2015) The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). Genome Biol Evol 7:3443–3462

Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol 13:1512–1517

Rokas A, Holland PWH (2000) Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15:454–459

Rota-Stabelli OR-SO, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, Pisani D, Philippe H, Telford MJ (2011) A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. Proc R Soc Lond B Biol Sci 278:298–306

Roy SW (2016) How common is parallel intron gain? Rapid evolution versus independent creation in recently created introns in daphnia. Mol Biol Evol 33:1902–1906

Roy SW, Gilbert W (2005a) Rates of intron loss and gain: implications for early eukaryotic evolution. Proc Natl Acad Sci U S A 102:5773–5778

Roy SW, Gilbert W (2005b) Resolution of a deep animal divergence by the pattern of intron conservation. Proc Natl Acad Sci U S A 102:4403–4408

Sankoff D, Leduc G, Antoine N, Paquin B, Lang BF, Cedergren R (1992) Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc Natl Acad Sci U S A 89:6575–6579

Schultz DW, Yarus M (1994) Transfer RNA mutation and the malleability of the genetic code. J Mol Biol 235:1377–1380

Sempere LF, Cole CN, McPeek MA, Peterson KJ (2006) The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. J Exp Zool Part B 306B:575–588

Sengupta S, Yang X, Higgs PG (2007) The mechanisms of codon reassignments in mitochondrial genetic codes. J Mol Evol 64:662–688

Shabalina SA, Koonin EV (2008) Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol 23:578–587

Shao R, Barker SC (2003) The highly rearranged mitochondrial genome of the plague thrips, Thrips imaginis (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA Genes. Mol Biol Evol 20:362–370

Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular systematics. BioEssays 22:148–160

Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N (1997) Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388:666–670

Skalsky RL, Cullen BR (2010) Viruses, microRNAs, and host interactions. Annu Rev Microbiol 64:123–141

Sperling EA, Pisani D, Peterson KJ (2011) Molecular paleobiological insights into the origin of the Brachiopoda. Evol Dev 13:290–303

Sperling EA, Vinther J, Moy VN, Wheeler BM, Semon M, Briggs DEG, Peterson KJ (2009) MicroRNAs resolve an apparent conflict between annelid systematics and their fossil record. Proc R Soc Lond B Biol Sci 276:4315–4322

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Sturtevant AH, Dobzhansky T (1936) Inversions in the third chromosome of wild races of *Drosophila pseudoobscura*, and their use in the study of the history of the species. Proc Natl Acad Sci U S A 22:448–450

Suh A, Paus M, Kiefmann M, Churakov G, Franke FA, Brosius J, Kriegs JO, Schmitz J (2011) Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. Nat Commun 2:443

Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biol 13:e1002224

Sverdlov AV, Rogozin IB, Babenko VN, Koonin EV (2005) Conservation versus parallel gains in intron evolution. Nucleic Acids Res 33:1741–1748

Takahashi K, Terai Y, Nishida M, Okada N (2001) Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Mol Biol Evol 18:2057–2066

Tanzer A, Amemiya CT, Kim CB, Stadler PF (2005) Evolution of microRNAs located within Hox gene clusters. J Exp Zool Part B 304B:75–85

Tarailo-Graovac M, Chen N (2009) Using repeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics Chapter 4, Unit 4, p 108

Tarver JE, Sperling EA, Nailor A, Heimberg AM, Robinson JM, King BL, Pisani D, Donoghue PCJ, Peterson KJ (2013) miRNAs: Small genes with big potential in metazoan phylogenetics. Mol Biol Evol 30:2369–2382

Telford MJ, Herniou EA, Russell RB, Littlewood DTJ (2000) Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. Proc Natl Acad Sci U S A 97:11359–11364

Thomson RC, Plachetzki DC, Mahler DL, Moore BR (2014) A critical appraisal of the use of microRNA data in phylogenetics. Proc Natl Acad Sci U S A 111:E3659–E3668

Venkatesh B, Ning Y, Brenner S (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. Proc Natl Acad Sci U S A 96:10267–10271

Waddell PJ, Kishino H, Ota R (2001) A phylogenetic foundation for comparative mammalian genomics. Genome Inform 12:141–154

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Yaakov B, Ceylan E, Domb K, Kashkush K (2012) Marker utility of miniature inverted-repeat transposable elements for wheat biodiversity and evolution. Theor Appl Genet 124:1365–1373

Yancopoulos S, Attie O, Friedberg R (2005) Efficient sorting of genomic permutations by translocation, inversion and block interchange. Bioinformatics 21:3340–3346

Zheng J, Rogozin IB, Koonin EV, Przytycka TM (2007) Support for the Coelomata Clade of Animals from a Rigorous Analysis of the Pattern of Intron Conservation. Mol Biol Evol 24:2583–2592