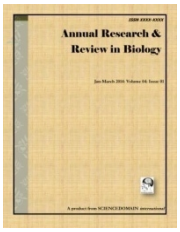

Annual Research & Review in Biology
4(10): 1518-1543, 2014

SCIENCEDOMAIN *international*
www.sciencedomain.org

Phylogenetic Networks: A Review of Methods to Display Evolutionary History

David A. Morrison[1*]

[1]*Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, 75189 Uppsala, Sweden.*

Author's contribution

The author DAM conceived the idea and wrote, read and approved the manuscript.

Review Article

Received 5th December 2013
Accepted 14th January 2014
Published 30th January 2014

ABSTRACT

Phylogenetic analysis attempts to reconstruct the genealogical history of evolutionary change in biological organisms. If the genealogy is complex, involving so-called horizontal evolutionary processes (such as recombination, hybridization, introgression and horizontal gene transfer) then an evolutionary network is required in order to graphically represent the history. Empirical examples of such networks have been used since the 1750s but only rarely. They fell out of favor from the late 1800s, when phylogenetic trees, which can represent only so-called vertical evolutionary processes (transfer of hereditary information directly from parent to offspring), were introduced to represent the Tree of Life. However, in the past 20 years there has been increased interest in using networks, as the evolutionary importance of horizontal processes has become increasingly more apparent. Unfortunately, there are currently few automated methods available, although this is an area of active algorithmic development. In this review, I discuss the development of both trees and networks as icons (or metaphors) for displaying phylogenetic relationships, to clarify some misunderstandings. I then provide an overview of the current approaches to using networks for the study of reticulate evolutionary relationships, explaining how the reticulation processes are detected based on the genetic patterns (or fingerprints) they produce. Finally, I review the current empirical use of evolutionary networks for displaying reticulate evolutionary histories. Due to the limitations of the current methods, many empirical networks have been produced manually or by modifying the output of a computer program.

**Corresponding author: Email: David.Morrison@slu.se;*

Keywords: Phylogenetics; evolution; genealogical relationships; phylogenetic networks; evolutionary networks.

1. INTRODUCTION

Phylogenetics tries to reconstruct the history of evolutionary change in biology. So, it is the study of the genealogical connection of all of life. Evolution involves a series of unobservable historical events, each of which is unique and we can neither make direct observations of them nor perform experiments to investigate them. This makes a phylogenetic study one of the hardest forms of data analysis known, as there is no mathematical algorithm for discovering unique historical accidents. This makes phylogenetics an interesting and challenging scientific discipline.

Reconstructing the evolutionary history of a group of contemporary organisms, called a phylogeny (or genealogy) is based on a study of their genotypic and phenotypic attributes. Each evolutionary event creates novel attributes and it is the pattern of shared attributes among the organisms that provides evidence of those events. Sources of evolutionary novelty include so-called vertical evolutionary processes, involving the passage of hereditary information directly from parent to offspring, including DNA substitutions, insertions and deletions, as well as genic duplications, translocations and inversions. Evolutionary novelties also come from so-called horizontal processes, such as recombination, hybridization, introgression, horizontal gene transfer and genome fusion, all of which transfer genetic information in more complex ways than by simple inheritance.

Historically, evolutionary history has been illustrated using the graphs known as either networks or trees. Phylogenetic trees are intended solely for the study of vertical evolutionary processes. Phylogenetic networks, on the other hand, are more general and can accommodate horizontal processes as well vertical ones. These horizontal processes are represented by reticulations in the network, which do not appear in a tree.

Mathematically, the trees and networks are connected graphs. These have labeled leaf nodes representing the contemporary organisms, internal nodes that are usually unlabeled, and edges connecting all of the nodes. Due to the complexity of evolutionary history, two types of graphs have been developed, which have been actively used in parallel by biologists for 250 years: (1) rooted evolutionary trees/networks (which are directed acyclicgraphs), in which the internal nodes represent ancestors of the leaf nodes and the directed edges represent historical pathways of transfer of genetic information between ancestors and their descendants and (2) unrooted affinity trees/networks (which are undirectedgraphs), in which the internal nodes do not represent ancestors and the undirected edges represent similarity relationships among the leaf nodes. In this review I am specifically concerned with the first type, the evolutionary diagrams, which explicitly display the genealogical history of the organisms (in the algorithmic literature, these have sometimes been called ‘explicit phylogenetic networks’); see the example in Fig. 1.

First, I discuss the development of both trees and networks as icons (or metaphors) for displaying phylogenetic relationships, to clarify some misunderstandings; then I provide an overview of the current network approaches to the study of reticulate evolutionary relationships (also called tokogenetic relationships) and finally I review the current use of evolutionary networks for displaying reticulate evolutionary histories.

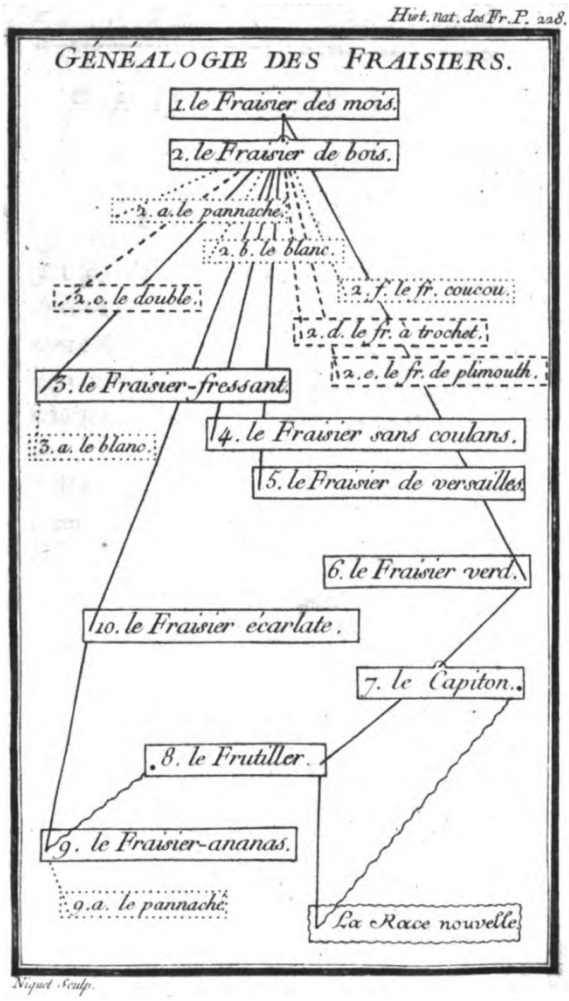

Fig. 1. The evolutionary network of 19 strawberry cultivars produced by Duchesne [16] (plate 1)

The root is at the top, with the genealogical history proceeding downwards. This is a hybridization network, with two hybrid cultivars indicated at the bottom

2. THE DEVELOPMENT OF TREES AND NETWORKS AS EVOLUTIONARY DIAGRAMS

The metaphors (or models) used for biological relationships have included a chain (or ladder), a tree and a network (or map) [1-3]. The chain was the accepted metaphor for most of recorded history, but the network (or web or map) was the predominant metaphor during the 1800s and the tree dominated during the 1900s [4-6]. However, this history also involves confusion between evolutionary and affinity diagrams and it is worthwhile clarifying this situation because there seems to be unnecessary confusion in recent discussions about the status of the so-called Tree of Life.

The first suggestion for replacing the chain was the network (Vitaliano Donati [7]), then the map (Carl Linnaeus [8]) and then the tree (Peter Simon Pallas [9]). Notably, the tree was explicitly intended as a simplification of the previously proposed network metaphor. That is, this was a form of parsimony analysis, in which the primary branches were considered more important and the reticulate relationships were treated as secondary. Moreover, all three metaphors (network, map, and tree) were based on affinity rather than genealogy.

Affinity refers to a natural overall group resemblance of organisms, usually quantified by some sort of weighted similarity of characters. It was the underlying concept of the Natural System of classification, which was such an important part of 18th and 19th century biology [10]. Patterns of affinity may, indeed, result from evolutionary relationships (i.e. genealogy), but affinity is a much broader concept than genealogy. In particular, affinity relationships are usually multi-directional rather than nested. In the modern world, affinity is usually represented by the “unrooted affinity” graphs referred to above.

None of these three people (Donati, Linnaeus, Pallas) actually included an illustrative fig with their metaphors, thus neatly separating theory and practice. However, networks of affinity did appear during the late 1700s, including those of Johann Philipp Rühling [11] on plants and Johann Hermann [12] on animals. The first published affinity map was actually Linnaeus’ own map of plant families, published by two of his former students, J.C. Fabricius and P.D. Giseke, in a posthumous collection of his lectures [13]. The first published affinity tree was that of Augustin Augier [14], which also concerned plant families. Interestingly, this temporal order of illustrations reflects the order in which the metaphors were originally proposed.

However, the first published evolutionary network was that of Georges-Louis Leclerc, Comte de Buffon [15], who illustrated the genealogy of dog breeds, which he postulated to include hybridization. This was followed by the work of Antoine Nicolas Duchesne [16], who studied hybridization among strawberry cultivars see Fig. 1. These figs pre-date any of the other diagrams, thus emphasizing the importance of reticulate relationships in the representation of evolutionary history. For the first published tree of genealogy we must wait until that of Jean-Baptiste Pierre Antoine de Monet, Chevalier de Lamarck in 1809 [17]. This was a transformational (or transmutational) tree rather than a modern phylogenetic tree, but it was nevertheless the first appearance of a dichotomous evolutionary diagram. For obvious reasons, there were no published maps of genealogy.

Affinity networks, however, were the dominant form of diagram during the first half of the 1800s, as evolution was not yet an accepted part of biological thought. Reticulating diagrams dominated over trees until the publication of Charles Robert Darwin’s [18] major work. Darwin’s ideas had two effects that are important for the discussion here. First, he focused attention solely on genealogical relationships, to the exclusion of all of the other relationships that are included in the concept of affinity. Second, he championed the tree as the appropriate metaphor.

Darwin [18] did not use the word ‘network’ but he did use the word ‘web’ with regard to affinity: “We can clearly see how it is that all living and extinct forms can be grouped together in one great system), and how the several members of each class are connected together by the most complex and radiating lines of affinities. We shall never, probably, disentangle the inextricable web of affinities between the members of any one class”.

An important point here is that affinity relationships can be expressed by an unrooted graph but genealogical relationships require a rooted (or directed) graph. That is a genealogy has a

time dimension, which is represented by directed edges pointing away from the root. So, in addition to the change from network to tree, there was an equally important change from undirected to directed metaphors. Indeed, Darwin [18] actually introduced the leap from affinity to genealogy via the tree metaphor. “The affinities of all the beings of the same class have sometimes been represented by a great tree. The green and budding twigs may represent existing species) and those produced during each former year may represent the long succession of extinct species.”

Darwin clearly named his ‘Tree of Life’ simile after its biblical namesake, and in doing so he “mobilized one of the oldest and richest traditions of imagery available to him. To play consciously on religious tree imagery was no new trick. but still it helped Darwin to seize the imagination of his readers” [19]. So, this simile was quite independent of Darwin’s famous bush-like diagram, because he always referred to his theory as “descent with modification” [20]. In this sense, Darwin’s motive for moving the metaphor from a network to a tree was use of a rhetorical device rather than being biologically motivated.

This distinction between different trees is important historically, because prior to Darwin the biblical tree imagery had already been co-opted to refer to the *arbor scientiae* (Tree of Knowledge), rather than the *arbor vitae* (Tree of Life). Nevertheless, Darwin’s ‘Tree of Life’ simile has since become widespread as a metaphor for phylogenetic relationships. These two Darwinian effects (the move to both genealogies and trees) were combined to reduce, for more than a century, the requirement for unrooted networks in biology and promote the use of rooted trees.

Consequently, rooted networks have mostly languished in biology since their appearance in the 1700s, with apparently only Ferdinand Albin Pax [21] producing them during the 1800s. In the literature from 1900–1990, I know of only nine evolutionary networks [22-30], two affinity networks, two networks representing convergence, and one network expressing genealogical uncertainty.

However, in the last 20 years biologists have seriously questioned this Darwinian change of metaphor in biology. In particular, there has been a concerted move to re-introduce the rooted network as the best metaphor for evolutionary history [31-35]. This move away from the tree is thus a return to where we were 250 years ago, rather than the introduction of a new metaphor. The past 150 years can thus be seen as a Darwinian diversion.

Along with this resurgence of interest, there has been interest in the development of new mathematical techniques of analysis [36,37]. Unfortunately, there are only a few such techniques currently available for evolutionary networks, and recent focus has therefore been on the development of practical and effective methods. This leaves contemporary biologists with a dilemma: the desire to use evolutionary networks but a lack of suitable techniques for producing them. Section 3 of this review discusses how biologists have been approaching the study of reticulate evolution under these circumstances, while Section 4 shows how evolutionary networks are presently being produced.

There is also some confusion in the literature about the appropriate terminology. Currently, ‘phylogenetic network’ is used indiscriminately to refer to both evolutionary networks and affinity networks. Indeed, the use of the word ‘net’ or ‘network’ to refer to evolutionary relationships is quite recent, as the authors in previous centuries used ‘arbre généalogique’ (French: family tree) or ‘verwandtschaftlichen’ (German: family relationship). (NB. This

terminology is quite natural, because a family pedigree is actually a network when both parents are included from each generation — each offspring is then a hybrid of two parents.) The first reference to a "phylogenetic net" was probably by Verne Grant in 1953 [25], who came close to modern usage when he used the expression to refer to what we would now call a 'hybridization network', which is a specific type of evolutionary network (as discussed in Section 4).

There was subsequently, however, a serious digression. The first reference to a "phylogenetic network" appears to have been by Holmquist [38]. Unfortunately, he used the word 'tree' to refer to a rooted tree and 'network' to refer to the unrooted topology of the tree. This usage was based on the idea that an unrooted tree represents a set of rooted trees, one potential root per edge in the tree. Avise et al. [39,40] then used the term "phylogenetic network" to refer to what is now called a 'haplotype network'. They manually created unrooted haplotype graphs by combining several gene networks. However, while there could be reticulations in their individual gene networks, the combined haplotype data had no reticulations, and they thus formed non-reticulate trees.

So, it was Bandelt [41] who formalized the reference to unrooted networks as 'phylogenetic networks'; previously, these had been simply called 'splits graphs'. This use of the term 'phylogenetic' does not refer to genealogies, since the graphs are unrooted and thus have no time dimension. That leaves open the question of who first used the expression 'phylogenetic network' explicitly in reference to a genealogy.

Finally, it is worth noting that the metaphor of a network is not the only possible one for reticulate relationships. For example, these alternative metaphors have been used for explicitly evolutionary relationships: Warp and weft [42], Intermingled blood streams [43], Trellis or lattice [44], Banyan tree [45], Rhizome [46], Braided river [47], Tree obscured by vines [48], Ring [49], Tree festooned with cobwebs [50] and Highways [51]. None of these metaphors seem to have been actively pursued.

3. RETICULATE PATTERNS AND PROCESSES IN EVOLUTIONARY HISTORY

Here, I provide an overview of the different processes and patterns involved in reticulate evolutionary history. I summarize the various evolutionary processes that will create reticulations in a genealogy, along with a brief description of the phylogenetic methods currently used to detect those processes in biological data and then construct a network. I do not cover simple inheritance from parent to offspring, because reconstructing a tree-like phylogenetic history is conceptually straightforward (although difficult in practice), and there are various existing reviews of that topic [52-54].

In phylogenetics, historical processes are expected to create patterns in contemporary species, and scientists then try to detect those patterns and assess them, in order to determine what process created each pattern. Computationally, algorithms will detect certain data patterns and display them in a directed acyclic graph (a tree or a network), which is then interpreted biologically. What is needed is to correctly identify the possible patterns created by the different processes, so that computational algorithms can be developed that will detect them. It is doubtful that an algorithm will be able to identify all of the individual processes — it will be up to biologists to work out what process created each of the detected patterns.

In this review I focus solely on using genotypes to study reticulate evolution. Traditionally, phenotypes have also been used (particularly to study hybridization), but this is now rarely done, except as a first hint that reticulation might be involved. The basic phylogenetic method for detecting reticulation processes based on their genetic data pattern is via incongruence involving gene trees [55] that is, a phylogenetic tree constructed for a single genic region. This incongruence can be used to detect hybridization, horizontal gene transfer and/or recombination. For example, the current mathematical algorithms [36] use incongruence between multiple gene trees to detect hybridization or horizontal gene transfer, and use incongruence within a single genome segment to detect recombination.

In what follows, there are major simplifications from both the biological and computational points of view. The reticulation processes are summarized in Table 1, along with the current methods of studying them. Literature examples of these analyses are discussed in Section 4.

Table 1. Summary of the patterns and processes involved in reticulate evolution

Reticulation process	Evaluation method
Polyploid hybridization (species)	multi-labeled tree
Homoploid hybridization (species)	incongruent gene trees
Homoploid hybridization (population)	sequence additive polymorphisms
Introgression (population)	haplotype network admixture graph
Horizontal gene transfer (species)	incongruent gene trees incongruent gene/species trees
Intra-genic recombination	sequence break-points
Inter-genic recombination	incongruent gene trees
Reassortment (population)	haplotype network
Genome fusion	genome similarities

3.1 Hybridization (Hybrid Speciation)

Hybridization refers to the formation of a new species via sexual reproduction [56,57]. The new (hybrid) species has a genome that consists of equal amounts of genomic material from each of the two parental species. There are two basic forms that are of interest:

Homoploid Hybridization, in which one copy of the genome is inherited from each parent species (e.g. diploid parents create a diploid hybrid);
Polyploid Hybridization, in which multiple copies of the genome are inherited from each parent species (e.g. diploid parents create a polyploid hybrid).

These concepts are illustrated in Fig. 2.

The existence of polyploid hybridization is usually assessed by DNA sequencing of each copy of the genome in the hybrid species, and then treating each copy as a separate unit in the phylogenetic data analysis. This produces a multi-labeled genome tree (i.e. each species appears multiple times in the tree, once for each copy of the genome), which is then turned into a single-labeled species network (i.e. each species appears only once in the network) [58].

At the species level, homoploid hybridization is usually assessed by sequencing several genes in the hybrid species (often from both the nuclear and the non-nuclear genomes) and producing independent trees for each gene. The species network is then created by

resolving conflicts among the topologies of the gene trees — the incompatibilities among the gene trees are resolved by postulating one or more reticulations in the network [59]. This form of analysis assumes a data pattern that is very similar to that of HGT (see Section 3.3).

In population studies (i.e. within species), homoploid hybridization is usually assessed at the sequence level, using either dominant (e.g. RAPD, ISSR) or co-dominant (e.g. nuclear DNA) makers [60]. Hybrids are detected by additive polymorphisms, where the genetic variation in the hybrid is the sum of the variation in the two parents. This may ocur, for example, at some positions in the sequence alignment of multiple-copy nuclear genes, when they are called superimposed nucleotide additivity patterns, or SNAPs). These polymorphisms arise either from (i) the polyploid nature of the hybrids (there are multiple copies of each chromosome, each of which may have a gene copy from either parental species), or (ii) from multiple paralogous copies of the genes within each genome.

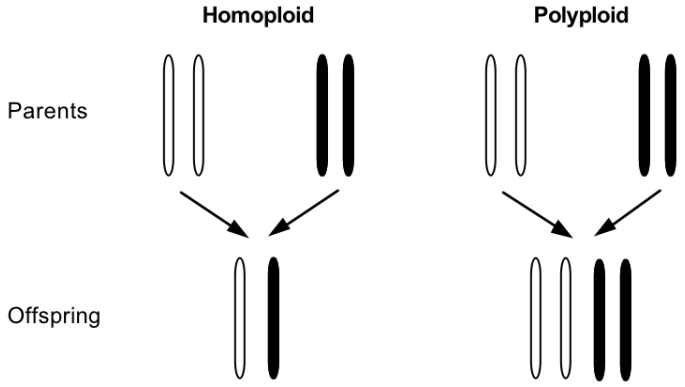

Fig. 2. Schematic representation of hybridization
The filled and open ovals represent chromosomes

3.2 Introgression (Introgressive Hybridization)

Introgression is the transfer of genetic material from one species to another via sexual reproduction [61]. This happens when hybrid individuals back-cross preferentially to one of the parental species, rather than forming a new hybrid species (as in Section 3.1). So, the genomes of one species are mosaics of genomic material from the two parental species, yet not necessarily with a 50:50 composition (at c. 50% it is best called hybridization rather than introgression). The data pattern created is very similar to that of HGT (see Section 3.3), as shown in Fig. 3.

Introgression is usually assessed at the population level, by sequencing one or more genes (often from both the nuclear and non-nuclear genomes) from many individuals, and demonstrating that identical haplotypes (haploid genotypes) are shared by what are recognized as separate species. This is done by constructing a haplotype network (as developed by [39,40]). Often, individuals are detected where the non-nuclear haplotype (usually inherited maternally) differs from the nuclear haplotype (inherited biparentally), as shown in Fig. 3. Note that the pattern produced by introgression is indistinguishable from that of heterogeneous genomic divergence (i.e. convergence and divergence are dynamic processes and they cannot be distinguished from a static pattern).

However, with the recent advent of the study of human genomes, single nucleotide polymorphism (or SNP) data are being used to detect admixture (another name for introgression) between different human populations (e.g. cultural or ethnic groups) and also between ancient humans and their close relatives (e.g. neandertals). The general approach is to add reticulations to an initial evolutionary tree constructed from the allele frequency data, so that the network fits the data better than does the tree [62].

At the species level, the distinction between introgression and hybrid speciation can be difficult to make. Indeed, discordance between organellar and nuclear phylogenies that appears to be ancient is perhaps more likely to be introgression rather than hybrid speciation [63]. In general, introgressive hybridization is probably much more common than hybrid speciation [64].

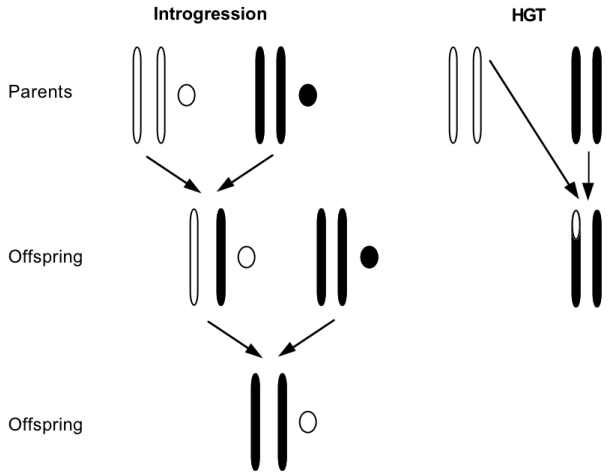

Fig. 3. Schematic representation of introgression and horizontal gene transfer
The filled and open ovals represent nuclear chromosomes and the filled and open circles represent organellar chromosomes

3.3 Horizontal Gene Transfer (Lateral Gene Transfer)

Horizontal Gene Transfer (HGT) is the transfer of genetic material from one species to another via non-sexual means (e.g. transformation, transduction, or conjugation) [65,66], as shown in Fig. 3. Thus, the genome of one species has some portion of its genomic material that was transferred from another (unrelated) species. The data pattern created is very similar to that of introgression (see Section 3.2), since the only essential difference between HGT and introgressive hybridization is that HGT does not occur via sexual reproduction while hybridization does. It is important to note, however, that HGT can potentially merge parts of distantly related species, whereas introgression can only combine closely related genomes.

HGT is sometimes assessed by sequencing several genes and producing independent gene trees. The species network is created by resolving incompatibilities among the gene trees, either the presence/absence of genes in some species or unexpected evolutionary relationships in some genes [67]. This is actually the same procedure as used for the study of homoploid hybridization, although hybridization involves whole genomes while HGT

usually involves partial genomes. Thus, this form of analysis assumes data that are very similar to those of homoploid hybridization or recombination.

Alternatively, HGT is often assessed by comparing observed gene trees to an expected species tree, which is either pre-specified or derived from multi-gene data [68]. The species network is then created by resolving conflicts between the gene trees and the species tree, particularly the existence of unexpected evolutionary relationships in some of the genes (especially the apparent genetic similarity of otherwise distantly related species).

3.4 Homologous Recombination and Viral Reassortment

Homologous recombination [69,70] and viral reassortment [71] are two processes that involve homologous parts of a genome breaking apart and re-arranging themselves. This usually occurs within a species, or between closely related species. In eukaryotes, it usually occurs during sexual reproduction. With crossing-over during meiosis the two genomes exchange material (called meiotic recombination), and with gene conversion one genome acquires material from the other.

There are three basic forms that are of interest:

Intra-genic Recombination, in which the break-points occur within a single gene;

Inter-genic Recombination, in which the break-points occur in different genes or non-coding spaces between genes;

Reassortment, in which segmented viruses re-combine their segments to create new strains (similar to gene conversion) when the two strains co-infect the same cell; this is basicallyinter-genic recombination without sex.

These are illustrated in Fig. 4. (Note that the expression “genetic recombination” is sometimes used more broadly than it is used here, to include any process that re-arranges bits of a genome, such as HGT or introgression.)

Intra-genic recombination is usually analyzed at the sequence level, based on ordered data such as the DNA sequences themselves. The gene network is constructed by identifying break-points, and thus the recombined segments [72]. It is also possible for one of the gene donors of a recombined sequence to be missing from the dataset, in which case the data pattern will be the same as for HGT (see Section 3.3) without the donor being sampled.

Inter-genic recombination can be studied similarly [73]. It will produce the same pattern as for homoploid hybridization (see Section 3.1) if both break-points are outside the region sequenced. Furthermore, homoploid hybridization can be thought of as recombination of whole chromosomes.

Viral reassortment is usually assessed by comparing the genomes of different virus strains with each other based on the presence/absence of segmental haplotypes (rather similar to haplotyping of sexual organisms) [74,75]. Incongruent phylogenetic relationships produced by different segments are used as evidence of reassortment [76]. This can produce incredibly complex networks [77].

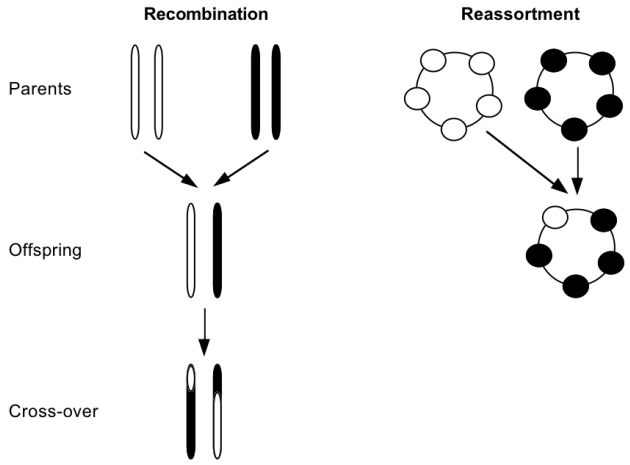

Fig. 4. Schematic representation of recombination and reassortment
The filled and open ovals represent chromosomes and the filled and open circles represent virus genome segments

3.5 Genome Fusion

Genome fusion (or symbiogenesis) is the addition of an entire genome of one species to that of another (unrelated) species by non-sexual means. Conceptually, it is an extreme form of HGT. It is usually considered to be rather rare, although it was presumably important in the origin of eukaryotes, mitochondria and chloroplasts, all of which appear to be fusions of the genomes of various prokaryotes [78,79].

This has always been assessed by comparing the genome in question with the genomes of other organisms, to identify genic similarities between them.

3.6 Distinguishing Between Processes Using Patterns

It may be possible to distinguish different types of evolutionary events, depending on the fingerprints or signatures that they leave in the data. If each cause of reticulation leaves a different data signature, then it is theoretically possible to devise models to detect those signatures and to then present a network representation of how they might have arisen.

Some explicit suggestions for recognizing signatures have been made, for example regarding the fingerprints of recombination versus hybridization [80]. Hybrid topologies can be topographically different from homologous recombination topologies because in the former the parental lineages remain unchanged, with the intermixed DNA in a new lineage. Consequently, there have been statistical tests suggested for detecting hybrids [81-83], which can be expected to have characteristics (phenotypic or genomic but not genic) intermediate between the parents.

However, in general, incongruence fingerprints are often identical among phenomena and it may be impossible ever to reliably distinguish homoploid hybridization, introgression, HGT and inter-genic recombination from each other by pattern analysis alone [84], at least not without genome-scale data. Moreover, the fingerprints of reticulation events may diminish

rapidly over time due to mutation or even other reticulation events; and reticulate evolution may be missed in studies using only a few genes [85].

Indeed, the only form of reticulate evolution for which we can currently reconstruct the history in a manner comparable to that for trees is polyploid hybridization. This is because there is an increase in chromosome number that is easily detected, and which thus acts as a clear fingerprint. HGT is sometimes considered to be reasonably obvious, because the end result can associate otherwise distantly related genomes, due to its non-sexual mechanism. However, there is some circularity to this argument.

Statistically, detecting a particular reticulation process involves a null tree model, so that rejecting the null model infers the presence of the process. However, any mathematical test is a test of the model not the process [86], so that rejecting the null model has no direct biological interpretation. Consequently, many of the network algorithms that have been developed are modeled on detecting a particular type of reticulation event but will end up detecting other types as well, rather than distinguishing between them.

Moreover, so far I have not discussed either deep coalescence (also called incomplete lineage sorting, or ancestral polymorphism) or gene duplication–loss which, if present, will confound the detection of reticulation patterns. Incomplete lineage sorting (ILS) and duplication–loss (D–L) are both vertical evolutionary processes, but they make the construction of a phylogenetic tree difficult, and they can confound the construction of a network by creating spurious reticulations. Deep coalescence (ILS) results in different gene fragments having different phylogenetic histories (i.e. the gene tree is not the same as the species tree), while gene duplication followed by selective loss (D–L) results in incomplete data for some gene fragments.

Therefore, an active area of research is the development of methods that will detect reticulation patterns in the presence of the potentially confounding effects of ILS and D–L. For example, there has been interest in detecting hybridization in the presence of ILS [87-94], HGT in the presence of ILS [95], HGT in the presence of D–L, called DLT [96-103] and HGT in the presence of both D–L and ILS [104].

4. CURRENT USAGE OF EVOLUTIONARY NETWORKS

In the previous Section, I discussed the detection of reticulate evolutionary histories. However, detecting a reticulate history does not imply that the history will necessarily be illustrated using an evolutionary network. There are many research papers where the focus is on estimating population parameters, for example, and studying the evolutionary processes but they do not present the species phylogeny itself.

So, most current research papers in evolutionary biology still do not illustrate reticulate evolution using a network genealogy. Instead, a collection of ad hoc methods is usually applied to the data, often involving a disparate array of software tools, and the evolutionary processes are then inferred from this. The use of a network to illustrate the reticulate genealogy inferred from this procedure is frustratingly rather rare.

Moreover, even when evolutionary networks are being used they are sometimes not identified as such. For example, they might be called ‘DLT reconciliation graphs’, or ‘ancestral recombination graphs’, or sometimes even ‘phylogenetic trees’. However, any graph that purports to explicitly display a reticulated genealogy *is* an evolutionary network.

Finally, while there is a great need for rooted evolutionary networks in phylogenetics, there are very few automated (or even semi-automated) methods available for constructing them. So, much of the study of reticulate evolution is still at least partly a manual process. Indeed, Moody and Rieseberg [85] have suggested: “Methods for depicting reticulate evolution in phylogenetic trees utilizing a network approach and existing data are becoming more advanced. However, reticulate evolution can become intractable with high levels of incongruence among many gene trees and combined with lineage sorting. Strictly empirical methods still need to be explored”.

Apparently, method development needs to progress a lot further before biologists are likely to use the methods for empirical studies. There is definitely a need for integrated software packages that can both detect and display reticulate evolutionary histories.

4.1 Overview

For species-level studies, most phylogenetics papers simply present a set of incongruent gene trees, although some papers also illustrate either (i) the tree derived from the combined-gene data or (ii) a consensus tree with or without the conflicting relationships or (iii) a pair of cophylogeny trees. Occasionally, the hybrid origin of some of the species, for example, is illustrated but the putative parents are not connected in a reticulate phylogeny. Alternatively, HGT events are inferred and the donor and recipient species are identified, but they are not explicitly connected in a reticulate phylogeny.

Population-level studies often present unrooted haplotype networks, illustrating processes such as hybridization and introgression between closely related species, or the evolution of domesticated species. These haplotype networks represent the genetic relationships among the different haploid genotypes observed in a dataset. They are usually drawn unrooted, since the root location is often unknown and so a haplotype network is not a true evolutionary network. In those cases where a root is provided [105], the network is still not an evolutionary diagram, because the reticulations in the graph represent uncertainty rather than genealogy.

However, the widespread use of ad hoc methods does not mean that evolutionary networks are absent from the literature. In the following subsections I present a representative sample of empirical studies where a rooted network has been used to illustrate an inferred genealogy. This will provide a brief introduction to the relevant literature. The examples are grouped according to the evolutionary processes being studied (see Section 3). In each case I have also briefly indicated how the networks were constructed.

However, first I will start with an example of how things have previously been done, in the absence of automated methods. Kellogg et al. [106] presented a complex collection of plant species where there appeared to be multiple causes of evolutionary reticulation. Their first step was to use cytogenetics to identify those species that are polyploids (i.e. by counting their chromosomes) and to remove them from further analysis. The polyploid hybridization network was constructed manually and displayed separately. For the diploid species they then constructed gene trees (by computer) for several nuclear genes, and identified those species that had incongruent placements in those trees. A species network was then constructed manually from these gene trees by adding reticulations (probably representing homoploid hybridization) to the tree that was common to all three genes. This is the classic approach to identifying reticulation — resolving incongruence among a set of gene trees.

Finally, the chloroplast tree (constructed by computer) was mapped onto the species network by manually adding more reticulations, which were taken to represent introgression.

4.2 Homoploid Hybridization

Homoploid hybridization is commonly studied in the literature and phylogenetic networks do appear, although not frequently. However, the mathematical model that has often been suggested for hybridization networks [36] does not fit many of the datasets collected by biologists. The usual mathematical model involves incongruence between two or more trees for the same set of species, for example from different genes or genomes. On the other hand, the data produced by biologists often involve only a single nuclear gene, and hybrids are detected by additive polymorphisms at alignment positions within the study gene.

This means that it is often difficult to apply any of the currently proposed computer programs for constructing evolutionary networks, for the phylogenetic analysis of many of the empirical datasets used for detecting hybridization. In turn, this suggests that we may need to develop a different mathematical model for hybridization networks, one based on additive polymorphisms rather than on incongruent trees.

As examples of the automatic construction of a hybridization network, Dickerman [107] used the unreleased program HyperPars (see Fig. 5) while Baum et al. [108] used the program PIRN [109] and Pirie et al. [110] used the program SplitsTree [111]. All of these programs can create evolutionary networks directly. As a semi-automatic procedure, Russell et al. [112] first constructed a network using the program Dendroscope [113] but then manually adjusted it, while Blanco-Pastor et al. [114] first constructed a multi-labeled tree from three genes. Note that the root of the network is not clearly indicated by either [110,112].

Sang et al. [115] constructed a network manually from a set of gene trees. Note that their network is drawn in a rather unusual style for indicating hybridization. Finally, Fuertes Aguilar and Nieto Feliner [116] constructed their network manually from additive polymorphisms in ITS sequences, while Fehrer et al. [117] manually constructed theirs from additive polymorphisms in the related ETS sequences.

4.3 Polyploid Hybridization

Polyploid hybridization is probably the most likely type of empirical study to contain an evolutionary network, particularly in botany, as plants (including ferns) are commonly polyploid. This frequency is at least partly because there is a computer program, PADRE [118], to automate much of the work. The mathematical model involves the construction of a multi-labeled genome tree (with data for each copy of the genome) that is then turned into a single-labeled species network.

PADRE has been used to construct hybridization networks by, for example, Marcussen et al. [92] and Sessa et al. [119,120] (see Fig. 6). As a different semi-automatic approach, Russell et al. [112] first constructed a network using the program Dendroscope, but then manually adjusted it (note that their root is ambiguously placed).

Alternatively, networks are sometimes constructed manually from one or more gene trees; for example, by Haufler and Windham [121], Marhold & Lihová [122] and Marques et al. [123]. Interestingly, the network of [121] has multiple roots, as drawn.

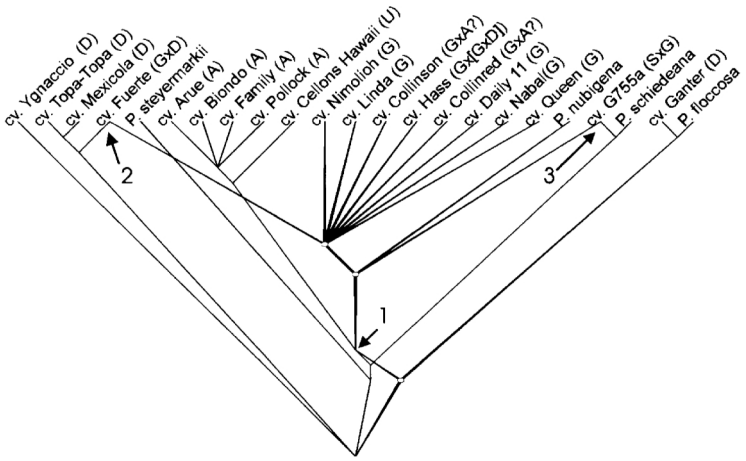

Fig. 5. The homoploid hybridization network of cultivars and wild species of avocados produced by Dickerman [107]

The root is at the bottom, with the genealogical history proceeding upwards. Three hybridization events are indicated by arrows. Reproduced with permission (from fig. 2)

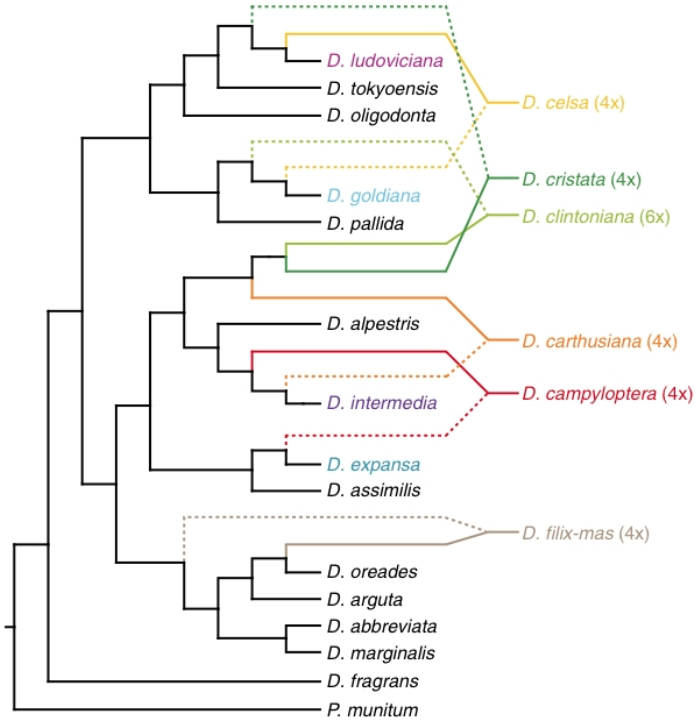

Fig. 6. The polyploid hybridization network of diploid and polyploid species of woodferns produced by Sessa et al. [119]

The root is at the left, with the genealogical history proceeding towards the right. Six hybridization events are indicated. Reproduced with permission (from Fig. 5)

4.4 Introgressive Hybridization

Introgression is a widely studied phenomenon. However, rooted evolutionary networks have rarely been presented. The usual mathematical model is to construct a haplotype network from data for haploid genotypes; but even a rooted haplotype network is not an evolutionary network.

Generally, an introgression network is constructed manually from a set of gene trees, as for example was done by Morgan [124] and Koblmüller et al. [125] (see Fig. 7). However, it is also possible to construct them automatically using programs designed to produce hybridization networks, such as SplitsTree, as was done by Labate and Robertson [126].

For the study of human admixture, several programs have recently been developed, including AdmixTools [127], TreeMix [128] and MixMapper [62]. These have been used to construct admixture networks from SNP data by, for example, Reich et al. [129,130]. Previously, these networks were constructed manually, as was also done by Reich et al. [131].

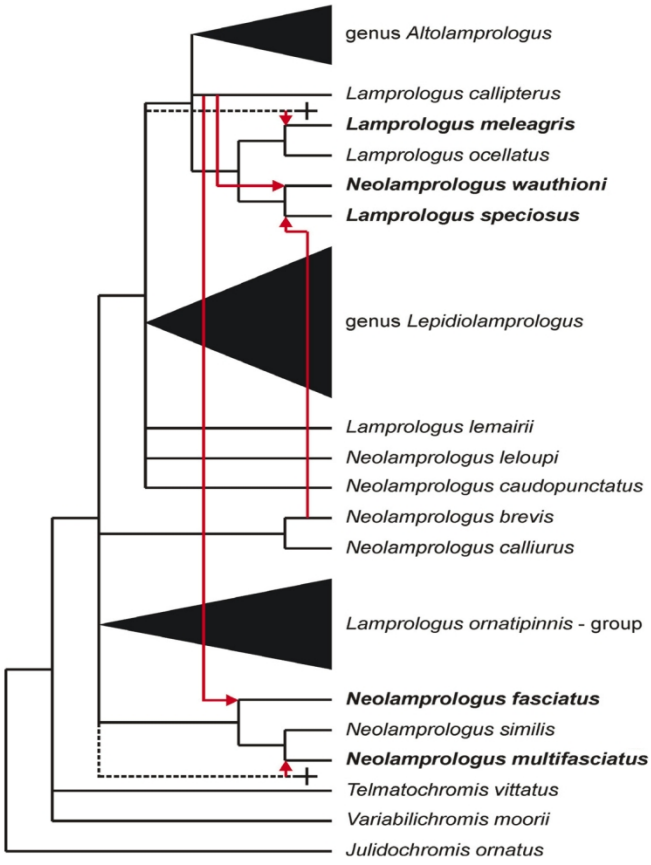

Fig. 7. The introgressive hybridization network of cichlid fish species produced by Koblmüller et al. [125]

The root is at the left, with the genealogical history proceeding towards the right. Five introgression events are indicated. Reproduced with permission (from Fig. 4)

4.5 Horizontal Gene Transfer

HGT is a hot topic these days, for both prokaryotes and eukaryotes (and even between them), although most research papers do not present a phylogenetic network. The usual mathematical model for lateral-transfer networks [36] is actually the same as one of those for hybridization networks, which does indeed seem to apply to the sort of datasets collected by biologists when they are studying HGT. That is, HGT is detected by incompatibility between two or more trees for the same set of species.

So, evolutionary networks do occasionally appear in the empirical literature. For example, Walsh et al. [132] automatically constructed a HGT network using the program SPRIT [133] based on incongruence between a species tree and a gene tree. On the other hand, Richards et al. [134] constructed their network manually from incongruence among a series of gene trees, as did Bergthorsson et al. [135] and Andersson et al. [136].

Furthermore, Delwiche and Palmer [137] (see Fig. 8) and Hao et al. [138] constructed their networks manually from a single gene tree that showed unexpected placements of various species. Using a single gene is not really recommended, however. For example, one of the single-gene datasets of Bergthorsson et al. [135] actually illustrates artifacts of the sequence alignment, rather than the HGT claimed by the authors.

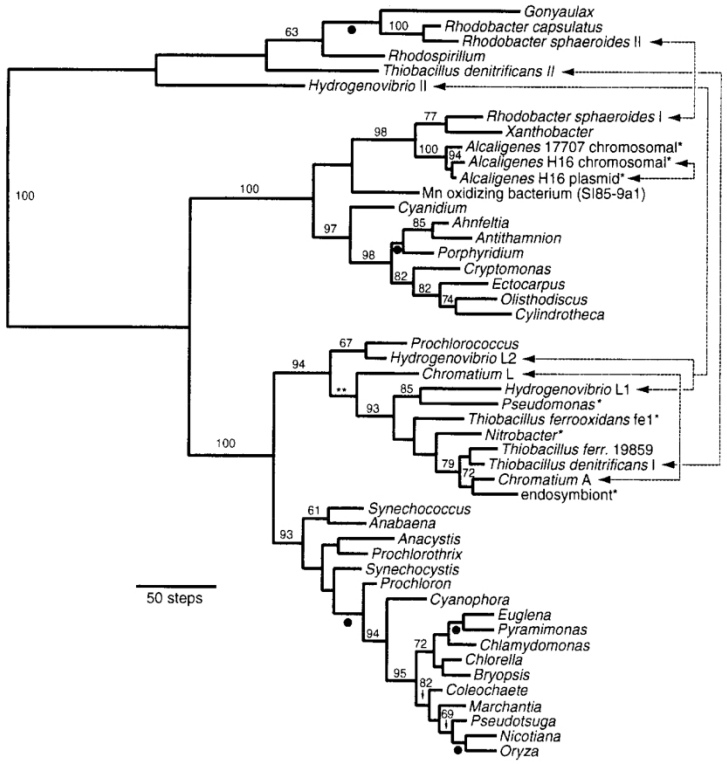

Fig. 8. The horizontal gene transfer network of bacteria and plastids (eukaryote organelles) produced by Delwiche & Palmer [137]

The root is at the left, with the genealogical history proceeding towards the right. Five HGT events are indicated. Reproduced with permission (from Fig. 2)

4.6 Homologous Recombination

Intra-genic recombination is often studied, but usually without reference to a network — the biological focus has been on detecting recombination rather than displaying the phylogenetic consequences. So, algorithmically, method development has also focused on detection of recombination; and these days recombination events are rarely detected manually.

Nevertheless, mathematical algorithms for recombination networks do exist [36], and there are several programs to automatically display the network. However, there seems to be little use of these programs outside of the algorithmic literature itself. In particular, the networks known as 'ancestral recombination graphs' are quite common in the algorithmic literature but are almost never seen in the empirical literature.

All the same, Jenkins et al. [139] constructed recombination networks using the program Kwarg [140] based on many different genes, Carbone et al. [141] (see Fig. 9) used the program Beagle [140] for their study of a gene cluster, and Morrell et al. [142] used the program SHRUB [143] to produce their set of recombination networks.

Chromosomal rearrangements are studied rather rarely, and there is little in the way of automated assistance. So, Rumpler et al. [144] constructed their network manually from a phylogenetic tree. Note that the root of their network is not clearly indicated.

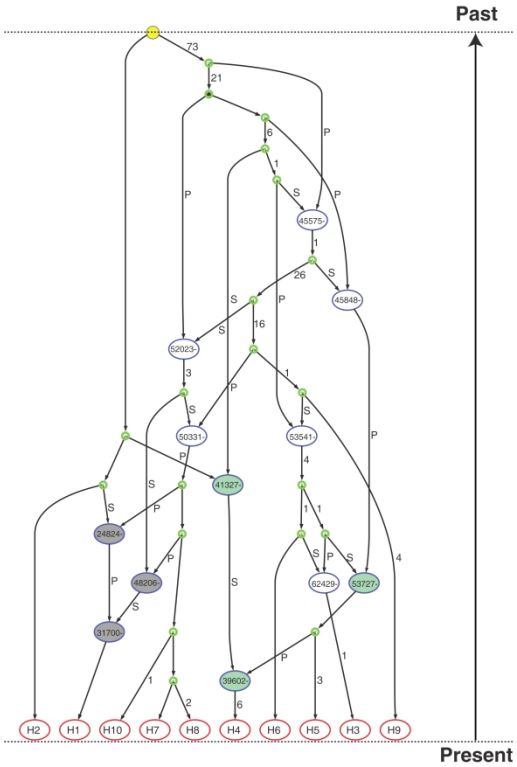

Fig. 9. The recombination network of fungus isolates produced by Carbone et al. [141]
The root is at the top, with the genealogical history proceeding downwards. Twelve recombination events are indicated. Reproduced with permission (from figure 2B)

4.7 Viral Reassortment

The reassortment of segmented viruses produces very complex networks. There can be innumerable possible paths through a reassortment network, the number increasing rapidly with the number of segments per virus. The networks are therefore usually constrained by the time at which a particular virus strain was first detected, so that each node in the network has a date.

Due to their complexity, only partial networks are usually seen in empirical studies, showing only those strains that are relevant to a particular viral outbreak (e.g. a pandemic). As an example, Smith et al. [145] manually constructed a partial network from a series of phylogenetic analyses.

Nevertheless, at least one automated method does exist [77] and Bokhari et al. [146] provide an example of its use. Many of the 'viral reassortment networks' published in the literature are unrooted [75] and therefore are not evolutionary networks.

4.8 Genome Fusion

Genome fusion is a difficult topic to study empirically; and most published evolutionary networks are theoretical rather than empirical. However, the network of Thiergart et al. [147] was constructed manually from a phylogenetic tree with unexpected placements of the various species and their component genomes.

4.9 Other Uses

There have been a few evolutionary networks published as part of empirical studies of topics not included above. I have included two examples here.

4.9.1 Apomixis

Apomixis is the study of non-sexual reproduction, which occurs not infrequently in plants, for example. This topic rarely involves the use of networks. However, an evolutionary network was constructed by Dyer et al. [148], who manually modified the output of the program SplitsTree.

4.9.2 Convergence

Evolutionary networks represent the history of particular evolutionary processes, but these processes do not have to involve actual reticulation events. For example, Alroy [149] used a hybridization network, constructed by an unreleased computer program, but interpreted the reticulations as being due to phenotypic convergence, rather than interpreting them as historical hybridization events. He noted that "the use of reticulations clarifies the phylogeny by factoring out apparent convergence, even though there is no reason to think that actual hybridization or introgression has occurred". That is, convergence will create apparent reticulation in an evolutionary history, rather than true reticulation, and it is up to the researcher to decide which is which.

5. CONCLUSION

Phylogenetic networks representing affinity are used to display conflicting phylogenetic signals. There are many different methods available, although the various forms of splits graphs seem to dominate. On the other hand, phylogenetic networks representing evolutionary history are less commonly found in the literature. There are few automated methods available and so most empirical networks are constructed either manually or by modifying the output of a computer program. This situation will hopefully change in the near future. In some ways this will return biologists to their original intention, 250 years ago, of explicitly depicting reticulate evolutionary histories. However, it will also move biology firmly into the 21st century, where the reconstruction of complex evolutionary scenarios can be considered to be a basic tool of all biological disciplines.

ACKNOWLEDGEMENTS

Thanks to Leo van Iersel, Steven Kelk and two anonymous referees for comments on an earlier version of the manuscript.

COMPETING INTERESTS

The author declares no competing interests.

REFERENCES

1. Kull K. Ladder, tree, web: the ages of biological understanding. Sign Syst Stud 2003;31:589-603.
2. Rieppel O. The series, the network and the tree: changing metaphors of order in nature. Biol Philos. 2010;25:475-496.
3. Pietsch TW. Trees of Life: a visual history of evolution. Baltimore: Johns Hopkins Uni Press; 2012.
4. Stevens PF. Metaphors and typology in the development of botanical systematics 1690–1960, or the art of putting new wine in old bottles. Taxon. 1984;33:169-211.
5. Barsanti G. Scala LA, mappa LA. l'albero: immagini e classificazioni della natura fra sei e ottocento. Firenze: Sansoni Editore; 1992.
6. Ragan MA. Trees and networks before and after Darwin. Biol Direct. 2009;4:43.
7. Donati V. Della storia naturale marina dell' Adriatico. Venezia: Francesco Storti; 1750.
8. Linnaeus C. Philosophia botanica. Amsterdam: Z. Chatelain; 1751.
9. Pallas PS. Elenchus zoophytorum. The Hague: Petrum van Cleef; 1776.
10. Stevens PF. The development of biological systematic. Antoine-Laurent de Jussieu, nature, and the natural system. Columbia Uni Press, New York; 1994.
11. Rühling JP. Ordines naturales plantarum commentatio botanica. Göttingen, Rosenbusch; 1774.
12. Hermann J. Tabula affinitatum animalium. Strasbourg, Treuttel; 1783.
13. Giseke PD. Caroli a Linné praelectiones in ordines naturales plantarum. Hamburg, B.G. Hoffmann; 1792.
14. Augier A. Essai d'une nouvelle classification des vegetaux. Lyon Bruyset Aine; 1801.
15. Buffon G-L de. Histoire naturelle générale et particulière, tome V. Paris Imprimerie Royale; 1755.
16. Duchesne AN. Histoire naturelle des fraisiers. Paris Didot le Jeune & C.J. Panckoucke; 1766.

17. Lamarck JBDE. Philosophie zoologique. Paris Dentu et l'Auteur; 1809.
18. Darwin C. On the origin of species by means of natural selection. London John Murray; 1859.
19. Hellström NP. The tree as evolutionary icon: TREE in the Natural History Museum, London. Arch Natural Hist. 2011;38:1-17.
20. Penny D. Darwin's theory of descent with modification, versus the biblical Tree of Life. PLoS Biol. 2011;9:1001096.
21. Pax FA. Monographische Übersicht über die Arten der Gattung *Primula*. Bot Jahrb Syst Pflanzeng Pflanzengeo. 1888;10:75-241.
22. Mereschkowsky C. Theorie der zwei Plasmaarten als Grundlage der Symbiogenese, einer neuen Lehre von der Entstehung der Organismen. Biolog Central. 1910;30:278-303,321-347,353-367.
23. Anderson E. Internal factors affecting discontinuity between species. Amer Nat. 1931;65:144-148.
24. Taylor H. Cyto-taxonomy and phylogeny of the Oleaceae. Brittonia. 1945;5:337-367.
25. Grant V. The role of hybridization in the evolution of the leafy-stemmed gillias. Evolution. 1953;7:51-64.
26. Goodspeed TH. The genus *Nicotiana*: origins, relationships and evolution of its species in the light of their distribution, morphology and cytogenetics. Chron Botan. 1954;16:1-536.
27. Lewis H, Lewis MRE. The genus *Clarkia*. Uni Calif Publ Bot. 1955;20:241-392.
28. Turner BL. A cytotaxonomic study of the genus *Hymenopappus* (Compositae). Rhodora. 1956;58:163-186,208-242,250-269,295-308.
29. Goodwin TW, Ochoa S. Comparative biochemistry of carotenoids. In Proceedings of the Fifth International Congress of Biochemistry, Moscow 10-16 Aug 1961, vol. III. Oxford, Pergamon Press; 1963.
30. Mikelsaar R. A view of early cellular evolution. J Mol Evol. 1987;25:168-183.
31. Doolittle WF. Phylogenetic classification and the Universal Tree Science. 1999;284:2124-2129.
32. Doolittle WF. The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. Phil Trans Roy Soc London B, Biol Sci. 2009;364:2221-2228.
33. Koonin EV. Darwinian evolution in the light of genomics. Nuc Acids Res. 2009;37:1011-1034.
34. O'Malley MA, Koonin EV. How stands the Tree of Life a century and a half after The Origin? Biol Direct. 2011;6:32.
35. Bapteste E, Van Iersel L, Janke A, Kelchner S, Kelk S, McInerney JO, Morrison DA, Nakhleh L, Steel M, Stougie L, Whitfield J. Networks expanding evolutionary thinking. Trends Genet. 2013;29:439-441.
36. Huson DH, Rupp R, Scornavacca C. Phylogenetic networks concepts, algorithms and applications. Cambridge: Cambridge Uni Press; 2011.
37. Morrison DA. Introduction to phylogenetic networks. Uppsala RJR Productions; 2011.
38. Holmquist R. A measure of the denseness of a phylogenetic network. J Mol Evol. 1978;11:225-231.
39. Avise JC, Giblin-Davidson C, Laerm J, Patton JC, Lansman RA. Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. Proc Natl Acad Sci USA. 1979;76:6694-6698.
40. Avise JC, Lansman RA, Shade RO. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I Population structure and evolution in the genus *Peromyscus*. Genetics. 1979;92:279-295.

41. Bandelt HJ. Phylogenetic networks. Verhand Naturwiss Vereins Hamburg. 1994;34:51-71.
42. Dixon RB. The building of cultures. New York: Charles Scribner's Sons; 1928.
43. Hooton EA. Up from the ape. New York: Macmillan; 1931.
44. Weidenreich F. Apes, giants and man. Chicago: Uni Chicago Press; 1946.
45. Linton R. The tree of culture. New York: Alfred A. Knopf; 1955.
46. Deleuze G, Guattari F. Mille plateaux. Les Editions de Minuit, Paris; 1980.
47. Moore JH. Putting anthropology back together again: the ethnogenetic critique of cladistic theory. Amer Anthropol. 1994;96:925-948.
48. Kim J, Salisbury BA. A tree obscured by vines: horizontal gene transfer and the median tree method of estimating species phylogeny. Pacific Symp Biocomput. 2001;6:571-582.
49. Rivera MC, Lake JA. The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature. 2004;431:182-185.
50. Ge F, Wang LS, Kim J. The cobweb of life revealed by genome-scale estimates of horizontal gene transfer. PLoS Biol. 2005;3:316.
51. Beiko RG, Harlow TJ, Ragan MA. Highways of gene sharing in prokaryotes. Proc Natl Acad Sci USA. 2005;102:14332-14337.
52. Semple C, Steel M. Phylogenetics. Oxford: Oxford Uni Press; 2003.
53. Felsenstein J. Inferring phylogenies. Sunderland (MA): Sinauer Associates; 2004.
54. Morrison DA. Phylogenetic analyses of parasites in the new millennium. Adv Parasitol. 2006;63:1-124.
55. Doyon JP, Ranwez V, Daubin V, Berry V. Models, algorithms and programs for phylogeny reconciliation. Brief Bioinform. 2011;12:392-400.
56. Arnold ML. Natural hybridization and evolution. New York: Oxford Uni Press; 1996.
57. Mallet J. Hybrid speciation. Nature. 2007;446:279-283.
58. Huber KT, Oxelman B, Lott M, Moulton V. Reconstructing the evolutionary history of polyploids from multilabeled trees. Mol Biol Evol. 2006;23:1784-1791.
59. Semple C, Gascuel O, Steel M. Hybridization networks. In Reconstructing evolution new mathematical and computational advances, Oxford Uni Press; 2007;277-314.
60. Whittall J, Liston A, Gisler S, Meinke RJ. Detecting nucleotide additivity from direct sequences is a SNAP: an example from *Sidalcea* (Malvaceae). Plant Biol. 2000;2:1-7.
61. Arnold ML. Evolution through genetic exchange. New York: Oxford Uni Press; 2006.
62. Lipson M, Loh PR, Levin A, Reich D, Patterson N and Berger B. Efficient moment-based inference of population admixture parameters and sources of gene flow. Mol Biol Evol. 2013;30:1788-1802.
63. Rieseberg LH. Hybrid origins of plant species. Ann Rev Ecol Syst. 1997;28:359-389.
64. Morgan-Richards M, Smissen RD, Shepherd LD, Wallis GP, Hayward JJ, Chan CH, et al. A review of genetic analyses of hybridisation in New Zealand. J Roy Soc NZ. 2009;39:15-34.
65. Boto L. Horizontal gene transfer in evolution facts and challenges. Proc Roy Soc London B, Biol Sci. 2010;277:819-827.
66. Gophna U. Lateral gene transfer in evolution. New York. Springer; 2013.
67. Poptsova M, Gogarten P. The power of phylogenetic approaches to detect horizontally transferred genes. BMC Evol Biol. 2007;7:45.
68. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nature Rev Genet. 2008;9:605-617.
69. San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. Ann Rev Biochem. 2008;77:229-257.
70. Wang WB, Jiang T, Gardner S. Detection of homologous recombination events in bacterial genomes. PLoS One. 2013;8:75230.

71. Nelson MI, Holmes EC. The evolution of epidemic influenza. Nat Genet. 2007;8:196-205.
72. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010;26:2462-2463.
73. Anderson CM, Chen SY, Dimon MT, Oke A, DeRisi JL, Fung JC. Recombine a suite of programs for detection and analysis of meiotic recombination in whole-genome datasets. PLoS One. 2011;6:25509.
74. Suzuki Y. A phylogenetic approach to detecting reassortments in viruses with segmented genomes Gene. 2010;464:11-16.
75. Svinti V, Cotton JA, McInerney JO. New approaches for unravelling reassortment pathways. BMC Evol Biol. 2013;13:1.
76. Nagarajan N, Kingsford C. GiRaF: robust, computational identification of influenza reassortments via graph mining. Nuc Acids Res. 2011;39:34.
77. Bokhari SH, Janies DA. Reassortment networks for investigating the evolution of segmented viruses. IEEE/ACM Trans. Comput Biol Bioinform. 2010;7:288-298.
78. Dyall SD, Brown MT, Johnson PJ. Ancient invasions: from endosymbionts to organelles. Science. 2004;304:253-257.
79. Duve C De. The origin of eukaryotes: a reappraisal. Nature Rev Genet. 2007;8:395-403.
80. Linder CR, Rieseberg LH. Reconstructing patterns of reticulate evolution in plants. Amer J Bot. 2004;91:1700-1708.
81. Sang T, Zhong Y. Testing hybridization hypotheses based on incongruent gene trees. Syst Biol. 2000;49:422-434.
82. Gauthier O, Lapointe FJ, Jajuga K, Sokolowski A, Bock HH. A comparison of alternative methods for detecting reticulation events in phylogenetic analysis. In Classification, clustering and data analysis: recent advances and applications Berlin Springer. 2002;341-347.
83. Gauthier O, Lapointe FJ. Hybrids and phylogenetics revisited a statistical test of hybridization using quartets. Syst Bot. 2007;32:8-15.
84. Holder MT, Anderson JA, Holloway AK. Difficulties in detecting hybridization. Syst Biol. 2001;50:978-982.
85. Moody ML, Rieseberg LH. Sorting through the chaff, nDNA gene trees for phylogenetic inference and hybrid identification of annual sunflowers (*Helianthus* sect. *Helianthus*). Mol Phylogenet Evol. 2012;64:145–155.
86. Joly S. JML: testing hybridization from species trees. Mol Ecol Res. 2012;12:179-184.
87. Meng C, Kubatko LS. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor Popul Biol. 2009;75:35-45.
88. Kubatko L. Identifying hybridization events in the presence of coalescence via model selection. Syst Biol. 2009;58:478-488.
89. Joly S, McLenachan PA, Lockhart PJ. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Amer Nat. 2009;174:54-70.
90. Bloomquist EW, Suchard MA. Unifying vertical and nonvertical evolution: a stochastic ARG-based framework. Syst Biol. 2010;59:27–41.
91. Yu Y, Than C, Degnan JH, Nakhleh L. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. Syst Biol. 2011;60:138-149.
92. Marcussen T, Jakobsen KS, Danihelka J, Ballard HE, Blaxland K, Brysting AK, Oxelman B. Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, Violaceae). Syst Biol. 2012;61:107-126.

93. Jones G, Sagitov S, Oxelman B. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. Syst Biol 2013;62:467-478.
94. Yu Y, Barnett RM, Nakhleh L. Parsimonious inference of hybridization in the presence of incomplete lineage sorting. Syst Biol. 2013;62:738-751.
95. Than C, Ruths D, Innan H, Nakhleh L. Confounding factors in HGT detection: statistical error, coalescent effects, and multiple solutions. J Comput Biol. 2007;14:517-535.
96. Mirkin BG, Fenner TI, Galperin MY, Koonin EV. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol. 2003;3:2.
97. Górecki P, Bourne PE, Gusfield D. Reconciliation problems for duplication, loss and horizontal gene transfer. In Proceedings of the 8th Annual International Conference on Research in Computational Molecular Biology, New York: ACM Press. 2004;316-325.
98. Hallett M, Lagergren J, Tofigh A, Bourne PE, Gusfield D. Simultaneous identification of duplications and lateral transfers. In Proceedings of the 8th Annual International Conference on Research in Computational Molecular Biology, New York: ACM Press. 2004;347-356.
99. Csürös M, Miklós I. A probabilistic model for gene content evolution with duplication, loss, and horizontal transfer. Lect Notes Comp Sci. 2006;3909:206-220.
100. Doyon JP, Scornavacca C, Gorbunov KY, Szöllösi GJ, Ranwez V, Berry V. An efficient algorithm for gene/species trees parsimonious reconciliation with losses, duplications and transfers. Lect Notes Comp Sci. 2010;6398:93-108.
101. Tofigh A, Hallett M, Lagergren J. Simultaneous identification of duplications and lateral gene transfers. IEEE/ACM Trans Comput Biol Bioinform. 2011;8:517-535.
102. Bansal MS, Alm EJ, Kellis M. Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. Bioinformatics. 2012;28:283-291.
103. Sjöstrand J, Sennblad B, Arvestad L, Lagergren J. DLRS: gene tree evolution in light of a species tree. Bioinformatics. 2012;28:2994-2995.
104. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics. 2012;28:i409–i415.
105. Witas HW, Tomczyk J, Jedrychowska-Danska K, Chaubey G, Ploszaj T. mtDNA from the Early Bronze Age to the Roman period suggests a genetic link between the Indian subcontinent and Mesopotamian cradle of civilization. PLoS One. 2013;8:73682.
106. Kellogg EA, Appels R, Mason-Gamer RJ. When genes tell different stories the diploid genera of Triticeae (Gramineae). Syst Bot. 1996;21:321-347.
107. Dickerman AW. Generalizing phylogenetic parsimony from the tree to the forest. Syst Biol. 1998;47:414-426.
108. Baum BR, Edwards T, Johnson DA. What does the 5S rRNA multigene family tell us about the origin of the annual Triticeae (Poaceae)? Genome. 2013;56:245-266.
109. Wu Y. Close lower and upper bounds for the minimum reticulate network of multiple phylogenetic trees. Bioinformatics. 2010;26:140-148.
110. Pirie MD, Humphreys AM, Barker NP, Linder HP. Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). Syst Biol. 2009;58:612-628.
111. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23:254-267.
112. Russell A, Samuel R, Klejna V, Barfuss MH, Rupp B, Chase MW. Reticulate evolution in diploid and tetraploid species of *Polystachya* (Orchidaceae) as shown by plastid DNA sequences and low-copy nuclear genes. Ann Bot. 2010;106:37-56.

113. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope an interactive viewer for large phylogenetic trees. BMC Bioinform. 2007;8:460.
114. Blanco-Pastor JL, Vargas P, Pfeil BE. Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. PLoS One. 2012;7:39089.
115. Sang T, Crawford D, Stuessy T. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Amer J Bot. 1997;84:1120-1136.
116. Fuertes Aguilar J, Nieto Feliner G. Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (*Armeria*, Plumbaginaceae). Mol Phylogen Evol. 2003;28:430-447.
117. Fehrer J, Krak K, Chrtek J. Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae) disentangling phylogenetic signal, reticulation, and noise. BMC Evol Biol. 2009;9:239.
118. Lott M, Spillner A, Huber KT, Moulton V. PADRE a package for analyzing and displaying reticulate evolution. Bioinformatics. 2009;25:1199-1200.
119. Sessa EB, Zimmer EA, Givnish TJ. Unraveling reticulate evolution in North American *Dryopteris* (Dryopteridaceae). BMC Evol Biol. 2012;12:104.
120. Sessa EB, Zimmer EA, Givnish TJ. Reticulate evolution on a global scale: a nuclear phylogeny for New World *Dryopteris* (Dryopteridaceae). Mol Phylogen Evol. 2012;64:563-581.
121. Haufler CH, Windham MD. New species of North American *Cystopteris* and *Polypodium*, with comments on their reticulate relationships. Amer Fern J. 1991;81:7-23.
122. Marhold K, Lihová J. Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. Plant Syst Evol. 2006;259:143-174.
123. Marques I, Feliner GN, Draper Munt D, Martins-Loução MA, Aguilar JF. Unraveling cryptic reticulate relationships and the origin of orphan hybrid disjunct populations in *Narcissus*. Evolution. 2010;64:2353-2368.
124. Morgan DR. nrDNA external transcribed spacer (ETS) sequence data, reticulate evolution, and the systematics of *Machaeranthera* (Asteraceae). Syst Bot. 2003;28:179-190.
125. Koblmüller S, Duftner N, Sefc KM, Aibara M, Stipacek M, Blanc M, et al. Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika — the result of repeated introgressive hybridization. BMC Evol Biol. 2007;7:7.
126. Labate JA, Robertson LD. Evidence of cryptic introgression in tomato (*Solanum lycopersicum* L.) based on wild tomato species alleles. BMC Plant Biol. 2012;12:133.
127. Patterson NJ, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. Genetics. 2012;192:1065-1093.
128. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012;8:1002967.
129. Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, et al. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Amer J Human Genet. 2011;89:516-528.
130. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al. Reconstructing Native American population history. Nature. 2012;488:370-374.
131. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009;461:489-494.
132. Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL. Widespread horizontal transfer of retrotransposons. Proc Natl Acad Sci USA. 2013;110:1012-1016.
133. Hill T, Nordström KJV, Thollesson M, Säfström TM, Vernersson AKE, Fredriksson R, Schiöth HB. SPRIT: identifying horizontal gene transfer in rooted phylogenetic trees. BMC Evol Biol. 2010;10:42.

134. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell. 2009;21:1897-1911.
135. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424:197-201.
136. Andersson JO, Hirt RP, Foster PG, Roger AJ. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. BMC Evol Biol. 2006;6:27.
137. Delwiche CF, Palmer JD. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol Biol Evol. 1996;13:873-882.
138. Hao W, Richardson AO, Zheng Y, Palmer JD. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proc Natl Acad Sci USA. 2010;107:21576-81.
139. Jenkins PA, Song YS, Brem RB. Genealogy-based methods for inference of historical recombination and gene flow and their application in *Saccharomyces cerevisiae.* PLoS One. 2012;7:46947.
140. Lyngsø RB, Song YS, Hein J. Minimum recombination histories by branch and bound. Lecture Notes in Computer Science. 2005;3692:239-250.
141. Carbone I, Jakobek JL, Ramirez-Prado JH, Horn BW. Recombination, balancing selection and adaptive evolution in the aflatoxin gene cluster of *Aspergillus parasiticus*. Mol Ecol. 2007;16:4401-4417.
142. Morrell PL, Toleno DM, Lundy KE, Clegg MT. Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. Genetics. 2006;173:1705-1723.
143. Song YS, Wu Y, Gusfield D. Efficient computation of close lower and upper bounds on the minimum number of recombinations in biological sequence evolution. Bioinformatics. 2005;21:I413-I422.
144. Rumpler Y, Hauwy M, Fausser JL, Roos C, Zaramody A, Andriaholinirina N, Zinner D. Comparing chromosomal and mitochondrial phylogenies of the Indriidae (Primates, Lemuriformes). Chrom Res. 2011;19:209-224.
145. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic Nature. 2009;459:1122-1125.
146. Bokhari SH, Pomeroy LW, Janies DA. Reassortment networks and the evolution of pandemic H1N1 swine-origin influenza. IEEE/ACM Trans Comput Biol Bioinform. 2012;9:214-227.
147. Thiergart T, Landan G, Schenk M, Dagan T, Martin WF. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. Genome Biol Evol. 2012;4:466-485.
148. Dyer RJ, Savolainen V, Schneider H. Apomixis and reticulate evolution in the *Asplenium monanthes* fern complex. Ann Bot. 2012;110:1515-1529.
149. Alroy J. Continuous track analysis: a new phylogenetic and biogeographic method. Syst Biol. 1995;44:152-178.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=419&id=32&aid=3478