

From basepairs to birdsongs: phylogenetic data in the age of genomics

John V. Freudenstein,^{a,*} Kurt M. Pickett,^b Mark P. Simmons,^{a,1} and John W. Wenzel^{b,2}

^a Herbarium, Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 1315 Kinnear Road, Columbus, OH 43212, USA

^b Department of Entomology, The Ohio State University, 1315 Kinnear Road, Columbus, OH 43212, USA

Accepted 19 May 2003

Abstract

Given the quantity of molecular data now available, including complete genomes for some organisms, one can ask whether there is a need for any data beyond complete genomic sequences for phylogenetic analysis. One reason to look beyond the genome is that not all character information is encoded in organismal genomes. We propose a hierarchy of characters that ranges from biologically transmitted but nongenomically encoded characters, such as bird songs, to characters that are genomically encoded. All of these characters can retain historical information and are potentially useful for phylogenetic analysis. In addition, a number of phenotypic levels that are expressions of the genome can be identified. The question whether it is worth including any of these levels if all of the underlying sequence data have been collected arises, since issues of redundancy occur. Utilization of phenotypic levels that are ultimately based on sequences may facilitate reconstructing homologies that are not evident from sequence data alone. We propose the use of simultaneous analysis of sequence data and as many levels of phenotypic characters as possible to take advantage of homology information that may be more easily recovered from the latter. A method that eliminates redundancy to the degree that it can be detected is proposed.

© 2003 The Willi Hennig Society. Published by Elsevier Inc. All rights reserved.

Genomics has been described as revolutionizing microbiology (Nierman et al., 2000), transforming biomedical research (Miller, 2000), and causing “an intellectual and experimental sea change” in biology as a whole (Vukmirovic and Tilghman, 2000). The first genome of a free-living organism was completed for *Haemophilus influenzae* in 1995 (Fleischmann et al., 1995). Since then, roughly one microbial genome has been completed every 2 months and the pace is expected to accelerate (Fraser et al., 2000; Nierman et al., 2000). About one vertebrate genome per year is expected to be completed (Miller, 2000). As of March 2003 on the EMBL website, complete genomes were available for 873 viruses, 308 organelles, 112 phages, 97 Bacteria, 16 Archaea, and

14 Eukaryota (lists of completed genomes are available through GenBank (<http://www.ncbi.nlm.nih.gov/>) and EMBL (www.ebi.ac.uk/genomes/)). Genomics is also revolutionizing phylogenetics (Brown, 1996). Phylogenetic analyses of entire genomes are commonly conducted for viruses (e.g., Lindstrom et al., 1998; Bollyky and Holmes, 1999; Vrati et al., 1999; Smith et al., 2000) and have been performed using the mitochondrial genome in chordates (Naylor and Brown, 1998), mammals (Allard et al., 1999), and birds and related vertebrates (Mindell et al., 1999). The availability of entire genomes does not in itself ensure satisfactory reconstruction of phylogenetic relationships. Establishing orthology among genes is often difficult for distantly related taxa. Sequence similarity, which generally is used alone (e.g., Makarova et al., 1999; Nelson et al., 1999), is insufficient to establish orthology and will often lead to misleading results (Thornton and DeSalle, 2000). Because only orthologous genes can be analyzed to infer phylogenetic relationships, only a minority of the genome may be suitable for phylogenetic analysis. For example, Huynen and Bork's

* Corresponding author. Fax: +614-292-3009.

E-mail address: freudenstein.1@osu.edu (J.V. Freudenstein).

¹ Present address: Department of Biology, Colorado State University, Fort Collins, CO 80523, USA.

² The order of authorship is alphabetical; all authors contributed equally.

(1998) analysis of the first 9 sequenced Archaea and Bacteria was limited to 34 genes. Nelson et al.'s (1999) analysis of 17 microbial genomes was restricted to 33 putative orthologs. Govan et al.'s (2000) analysis of positive-stranded RNA viruses was limited to the conserved RdRp domain of 118 amino acids.

The problem of orthology is particularly acute for viruses, the genomes of which are mostly shorter than 10 kb (e.g., Bollyky and Holmes's (1999) phylogenetic analysis of mammalian hepadnaviruses was based on genomes of about 3.2 kb). This is shorter than the 12,234 protein-coding nucleotide sites of the mitochondrial genome in chordates, which were found to support an "incorrect" tree, regardless of the character coding (nucleotide, base, or amino acid) or tree construction method (parsimony, distance, or maximum-likelihood) employed (Naylor and Brown, 1998). Analyzing the entire viral genome, Bollyky and Holmes (1999) recovered phylogenetic trees whose topologies were highly dependent on the maximum-likelihood model used. Likewise, using the entire mitochondrial genome of birds and relatives, Mindell et al. (1999) found that their ingroup was ambiguously rooted because the resolution depended on the maximum-likelihood model used.

If whole genomic sequences may be insufficient for phylogenetic analyses, then where are other characters to come from? If the information from morphology and other phenotypic characters has already been captured when whole genomes are analyzed, recommendations to combine data would seem to be at a loss when there is apparently nothing remaining to combine with the sequence data. We argue that there are two sources of characters, additional to the genome, that may be useful.

The first source is the phenotypic characters just mentioned. *Phenotype* in its traditional sense can be construed quite broadly—Johannsen's (1909) definition as the expression of a genotype in interaction with its environment is sometimes broadened even further to mean all of the characteristics of an individual. We focus here on the determining factor for *character states*—if the character state as expressed in the individual is determined by (i.e., can be completely predicted by examination of) the genome, then we consider that expression to be phenotypic. However, if the particular expression of a state is determined by some other agent (such as the environment), even though the basis for the character is in the genome, we would consider this to be nonphenotypic. Examples of phenotypic characters would be amino acids and morphological structures (to the extent that their final state is determined by the genome). We argue that such characters have a useful role to play in improving the reconstruction of character state transformations when analyzed in conjunction with genomic characters, as described below.

The second source of characters are those features of organisms not encoded by the genomes. Examples of such features are centrosomes, prions, and behaviors, even though, at least for the latter two, the ultimate basis for the character resides in the genome. The key is that the particular *state* exhibited is not encoded in the genome. In what follows we describe a way of viewing the different types of characters potentially useful for phylogenetic analysis and present a method that maximizes use of data.

Hierarchy of information

If we consider the range of possible phylogenetic indicators—those features that can store information relevant to reconstructing the history of a clade—the list is long and includes some elements that are not commonly used for this purpose. We conceive of these features as forming a hierarchic, nested set defined by a series of properties (depicted as a Venn diagram; Fig. 1). This hierarchy is neither perfect nor the only way to organize this information, but it is a useful structure in which to focus on the features that are important in the selection of phylogenetic markers. For each level we describe the defining property of that level and then give some examples that fall within that level but outside of the next lower level.

The most inclusive level comprises all features of a species and its environment. The key question is which of these features retain historical information about the species' relationships. Some features are intrinsic to the organisms while others are not. A specific example of the latter would be the geographic location in which a species occurs. To the extent that a species and its descendants remain in the same place, location becomes an historical attribute of a clade. However, this attribute is purely circumstantial, being nonintrinsic and inherently nonassessable from examination of specimens them-

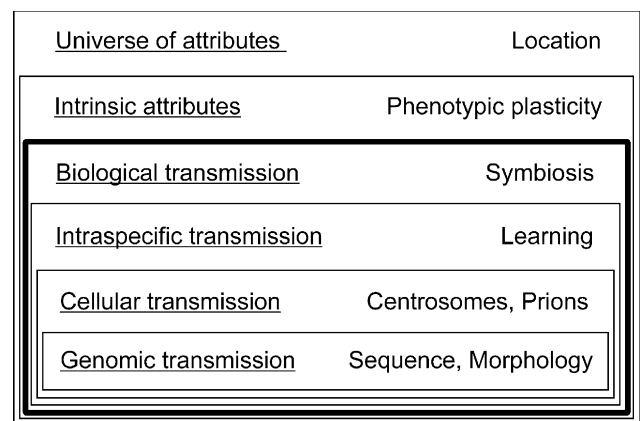


Fig. 1. Hierarchic depiction of organismal attributes that may be considered for phylogenetic analysis.

selves. Such biogeographic information is often mapped onto phylogenetic diagrams a posteriori, but rarely has been used in their construction (but see Dressler, 1990, p. 122).

A subset of all features are those that are intrinsic properties of the organism itself in some way. These include both biologically and nonbiologically determined features. An example of the latter would be variation due to phenotypic plasticity. Such features have a genetic basis, in that different phenotypes are possible depending upon the environment in which the organisms occur, corresponding to the classical concept of “norm of reaction.” A common example is the relationship of plant height to altitude, such as was shown for various species in the classic experiments of Clausen et al. (1940). To the extent that a species persists under particular environmental conditions, a specific phenotype will persist and retain historical information. However, if the environment in which a descendant species finds itself has changed relative to that of an ancestor and plasticity has been retained, the features of the descendant species could revert to an alternate phenotype. Hence, the specific phenotype need not be transmitted to descendants. This includes, for example, the feature of cell adhesion discussed in a hypothetical example by Newman and Müller (2000).

The remaining levels are sharply distinguished from the previous levels in that the agent of information transmission is biological. This is a crucial distinction because biologically transmitted attributes are more strongly associated with taxa than those that are nonbiologically transmitted or merely circumstantial. Hence, it is at this level that we speak of *characters*, which we define simply as biologically transmitted attributes of a species.

At the next level are systems in which only one species is involved in the transmission of the information that forms the character. Such characters include behaviors that are learned in each generation from preceding generations.

The next level delimits those features that are transmitted via cellular processes. These include structural features such as centrioles and prions, whose form is not determined by the genome, and features encoded by the genome. Although prions are ultimately encoded by the genome, their alternate conformations, which we would code as states, are not genetically determined (see below).

The final level includes those features that are encoded by any of the genomes of the organism. These include nuclear, mitochondrial, and plastid sequences and any phenotypic characters that are derived from those sequences, including morphological characters. This least inclusive class includes the great majority of characters that are commonly employed in current

phylogenetic analyses. In what follows we focus in more detail on some key levels.

Biological transmission

Phylogenetic study is oriented toward biological transmission of information among taxa and specifically to information encoded in the genome. The evolutionary biology community has a long-standing aversion to Lamarckian views on “inheritance of acquired characteristics,” because of their failure to provide a mechanism for inheritance. Accordingly, objection should disappear when a mechanism is evident. Mutations, after all, are simply acquired changes of the genome. DNA replication is the mechanism of genetic inheritance, but there are many other features of an organism that are not encoded in the genome, yet are replicated. Even Richard Dawkins (1976, pp. 191–192) has argued that replicators need not be genetic: “I am an enthusiastic Darwinian, but I think Darwinism is too big a theory to be confined to the narrow context of the gene. The gene will enter my thesis as an analogy, nothing more. What, after all, is so special about genes?”

Given the possibility of extragenomic replication, are there nongenomically encoded features that can still be inherited, show variation, and therefore serve as suitable characters for phylogenetic analysis? If there are such features, they would be missed in an analysis based solely on the genome. We argue that these characters exist and present examples below.

Many examples of biologically transmitted attributes that are not encoded in the genome can be described. One broad category would contain tightly linked symbioses, which themselves can give rise to useful phylogenetic characters. Well-known examples of symbioses include lichens composed of fungi and algae, termites harboring wood-digesting flagellates in their guts (Cleveland, 1926), ants that feed exclusively from fungus gardens that they grow (Chapela et al., 1994), and parasitic wasps that rely on polydnviruses to overwhelm the host immune system (Whitfield, 2000). Attributes of one of the partners can be used to inform us about the phylogenetic history of the other, just as with parasites. Indeed, features of parasites are sometimes used as characters of their hosts in systematic studies (e.g., Eichler, 1941; Brooks, 1981).

If the symbiotic association is very close, it may result in a dependence in which one or both partners cannot live without the other. For example, some leaf-cutter ants have apparently lost ordinary digestive enzymes because of their reliance upon fungi to process plant matter (Martin, 1987). In extreme cases, the association may produce novel features that are something more than the sum of the parts. For example, lichen relationships are known to result in the production of novel chemical substances not present when the members of

the association are free-living (see review in Brodo et al., 2001, pp. 42–43). Protracted dependency and synergism may result in such well-known relationships as eukaryotic cells containing mitochondria and plastids. This demonstrates that symbiosis can lead eventually to our most fundamental character level, genomic transmission (Fig. 1). Interesting intermediates may exist for which placement in our proposed hierarchy is not immediately clear. An example is the bacterium *Wolbachia*, which is transmitted in the cells of other taxa while remaining an independent organism (Yen and Barr, 1971).

In some cases, details about characteristics of location, such as “lives at deep sea vents” for a kind of crustacean, serve as plausible surrogates for heritable characters of physiology, whereas alternative descriptions are less clearly related to informative biological aspects (Miller and Wenzel, 1995). Perhaps “lives at the mid-Atlantic ridge” is also correct, but is meaningful geographically rather than biologically. The distinction between these cases may be vague, and perhaps a good rule of thumb is whether the organism would be expected to survive if it were kept in a similar but alternative environment. If the crustacean would not be expected to survive in the Atlantic benthos away from the vent, then “lives at deep sea vents” seems to be a useful biological characteristic. By contrast, kangaroos do not cease to be kangaroos in any meaningful way when they are removed from Australia. This can be considered a question of *sine qua non*; if the property in question is essential and indispensable to the individuals of a species, then it surely meets the criterion that we use for phylogenetic characters. The criterion can apply to other characters, such as those of symbiosis: If depriving a termite of its symbionts causes the termite to cease vital function, then the symbionts can be considered characteristics of the termite. Not all individuals of a species will necessarily bear parasites at all times that are characteristic of the species—thus, the criterion of indispensability is sufficient, but not necessary.

Intraspecific transmission

The boundary between intrinsic and extrinsic influences can be indistinct when learning plays a role. A continuum seems to exist from situations where genetic determinism is a reasonable assumption to those where transmission of information from one generation to the next is clearly extragenetic. For example, animals may have a genetic predilection for performing a kind of behavior and a genetic template or rule-of-thumb for evaluating correct form, but nonetheless the behavior itself is shaped through repeated trials and progressive learning of the skills necessary to complete the desired product. The first and last steps are predetermined and the animal simply fills in the intermediate links, each individual learning to do so independently. An example

of this might be nest building in weaver birds who learn to tie the knots that are necessary to make their stereotypical nests (Collias and Collias, 1984). The first step toward extragenetic transmission comes when birds that observe other weavers learn to build their own nests more rapidly, perhaps in 4 months rather than 5 (Collias and Collias, 1984, p. 222). The ability to learn from other individuals permits a cultural lineage to provide historical data independent of genetic data.

Perhaps the best example of persistent culture outside of humans is found in bird song. In some sparrows, songs are learned and consolidated by young individuals, and then the song remains unchanged throughout the birds' adult life and experience (Marler and Tamura, 1964; Marler and Peters, 1981). Call learning may permit incorporation of elements not just from other individuals, but even from other species (Gaunt et al., 1994). Mundinger (1979) specifically discussed eliminating learned elements to find phylogenetically informative elements, although he also concluded that learning itself is a useful character in finches. More critically, studies of various groups have found repeatedly that persistent local dialects are not related to genetic structure of the populations and therefore that the distinction between dialects themselves is not likely to be genetic. This has been shown in sparrows (Lougheed and Handford, 1992), cowbirds (Fleischer and Rothstein, 1988), and parrots (Wright and Wilkinson, 2001) and is inferred in hummingbirds (S.L. Gaunt, pers. comm., 2001). These authors and others (e.g., Zink and Barrowclough, 1984; Zink, 1985) comment specifically that the call dialects are conserved in the face of high gene flow across dialect boundaries. In other species, genetic distinctions are found at some but not all dialect boundaries (Kroodsma et al., 1985; Balaban, 1988), one possible explanation being that the dialects form first, and genetic distinctions between populations sometimes follow. Indeed, Wright and Wilkinson (2001) compared closely related species and deduced that “propensity to form [temporally stable] dialects can be inferred to have been present in the common ancestor of this clade.”

Bird song presents us with an example where there is vertical transmission of information extrinsic to the genome, where variation in this information can be more stable than the genotypes of the individuals themselves, and where subsequent genetic variation might be structured according to the extrinsic, extragenetic attributes. Thus, it is possible that some variation in external attributes precedes genetic variation, the exact opposite of what is generally assumed. The examples used here span five families in three orders [Emberizidae, Fringillidae, and Icteridae, (Passiformes); Trochilidae (Apodiformes); Psittacidae (Psittaciformes)], so this phenomenon is not restricted to certain peculiar lineages. Whether this type of vertical transmission in learning is found in other systems will not be known until other systems are

as thoroughly explored as bird songs are, but we can expect to find analogies. For example, host choice may be one, and it is already known from cross-fostering experiments that certain parasitic wasps favor the hosts that their mothers chose (Turlings et al., 1993). The same may apply to nest site selection in wasps (Wenzel, 1996). Perhaps animal migration will be similar, as suggested by the situation known widely among laymen of a researcher flying an ultra-light airplane to direct eager flocks of naïve birds. The animals inherit an instinct to migrate but the details of migration, which may remain stable over evolutionary time, are nonetheless transmitted biologically and extragenetically.

We realize that some may find the use of these characters controversial. One common claim is that the “character tree” resulting from certain kinds of character data will not be the same as the “taxon tree” (see review in Doyle, 1991). We assert that all phylogenetic hypotheses are character phylogenies and that the validity of any character or suite of characters is best evaluated in simultaneous analysis with other characters (Kluge, 1989; Doyle, 1991; Nixon and Carpenter, 1996).

Cellular transmission

Centrosomes, and the mechanism by which they are inherited, have been called “a central enigma of cell biology” (Wheatley, 1982). This enigma stems from the centrosome’s function in the cell and its apparent autotransmission. Centrosomes are organelles composed of two paired centrioles surrounded by the dense, amorphous pericentriolar material. The centrosome plays an important role in maintaining the structure of the cell by generating the kinetochore fibers of the mitotic spindle apparatus (Nicklas, 1971; Wheatley, 1982), which are responsible for chromosome movement during metaphase (Nicklas and Koch, 1972) and cell cleavage (Rappaport, 1986).

Centrosomes are also responsible for a number of phylogenetically useful traits, including the creation of retinal rods, the motility of epithelial cells, and proper antigen reorganization in lymphocytes (Brown et al., 1992). For a cell to form a cilium, it must have at least one centrosome; protistologists have long used the presence of cilia as a diagnostic character in their phylogenetic studies (Corliss, 1979).

In biparental, diploid animal species, centrosomes are disassembled in both male and female gametes and differentially reassemble at the time of fertilization: sperm-derived centrioles are assembled with egg-derived pericentriolar material to form the daughter centrosome (Schatten, 1994; Callaini et al., 1999; Palazzo et al., 2000). In haplodiploid species, centrosome inheritance is maternal when males are formed and paternal when females are formed (Tram and Sullivan, 2000). The mechanisms driving these inheritance phenomena are

unknown. Although at least one centrosome protein, centrosomin, has been shown to derive from nuclear DNA, when this protein is mutated in *Drosophila*, the centrosome organizes normally, but the individual’s sperm are without flagella and so are rendered ineffective (Li et al., 1998). It is not clear, however, whether all centrosome proteins are produced by the cell. Indeed, as Sluder (1992, p. 254) stated, strong evidence suggests that this is not the case: “Neither transcription, translation, nor nuclear DNA synthesis are required for the repeated reproduction of sperm centrosome.” The inherited protein components appear to be used in the replication of a new centriole, and when the necessary components are present *in vitro*, protein microtubules are formed (Weisenberg, 1972), although centrioles themselves are not, implying that another centriole need be present to serve as an information-bearing template. At a minimum, critical inquiry has failed to demonstrate the redundancy of phenotypes and genotype in the case of centrosomes, though not for want of trying. Therefore, inclusion of these extragenetic characters in a simultaneous analysis with characters that are genomically transmitted is warranted.

Another source of characters that are not genomically transmitted is prions. The term *prion* originally described a particular kind of cellular protein that is the etiologic agent of the transmissible spongiform encephalopathies (Prusiner, 1994; Weissmann, 1994). However, most identified prions are not disease causing, but are functional (for a review of these, see Cox, 1965; Lacroute, 1971; Wickner, 1994; Tuite and Lindquist, 1996). These proteins are curious for at least five reasons: they can alter their own conformation in many ways, resulting in multiple types that evoke new phenotypes in the organism (variation; Parchi et al., 1996; Collinge et al., 1996); the resulting types (or novel character states) can confer special properties on the organism that alter (enhance) its survivability (fitness; Magasanik, 1992; Hofstetter et al., 1974; Lindquist et al., 1995); they can alter surrounding proteins, converting them to prions (within-organism transformation; Prusiner, 1991; Weissmann, 1994); they are transmissible between organisms (heritability; Cox, 1965; Lacroute, 1971; Tuite and Lindquist, 1996), and once formed, they change without a change in nucleic acid sequence (nongenetic inheritance; Lindquist et al., 2001; Serio and Lindquist, 2000; Tuite, 2000). Additionally, the alternate conformational states of prions (Parchi et al., 1996; Collinge et al., 1996; Lindquist et al., 2001) are maintained across generations in yeast (Sondheimer et al., 2001), showing no evidence of reversion to the wild-type form. This fixation is an important property of any replicator if it is to reflect phylogeny. If prions change randomly from wild-type to mutant and back again rapidly within a population, we might not expect the generation of synapomorphy.

Prions exhibit heritable variation, potentially generating synapomorphy and marking ancient divergences. However, the primary protein structure of prions does not encode their conformational structure; this means that not all of their heritable information is encoded in genomes and that this variation will need to be coded separately.

Genomic transmission

Genomic sequences have become well established as a source of phylogenetic information; the preponderance of systematic studies employing these data speaks to their importance. Phenotypic characters are encoded by the genome, but reflect underlying genomic changes to a greater or lesser degree. The most familiar phenotypic character type is morphology, which was the first source of systematic data for most taxa. Some authors have questioned the usefulness of morphological data for phylogenetic analysis as compared to molecular data (e.g., Sibley and Ahlquist, 1987; Gottlieb, 1988; Sytsma et al., 1991; Graur, 1993; Hedges and Sibley, 1994; Hedges and Maxson, 1996; Givnish and Sytsma, 1997a,b). The argument essentially reduces to one of ability to assess homology (the perception that molecular data have less homoplasy than morphological data), but homology hypotheses clearly can be problematic with both types of data, as witnessed by the problem of sequence alignment (Gatesy et al., 1993; Lutzoni et al., 2000). While Hillis (1987) suggested that at deep levels it might be more difficult to homologize morphological characters than molecular ones, Lanyon (1988) argued the opposite, indicating that at least in some cases, morphological characters may be more easily reconstructed as synapomorphies than more variable molecular states. Philippe and Adoutte (1998) suggested that molecular sequences might be insufficient to resolve eukaryote phylogeny and argued for careful selection of morphological and biochemical characters. Examples of analyses in which combined morphological and molecular datasets yield better-supported trees than either dataset alone are common (e.g., Freudenstein, 1999; Simmons et al., 2001), suggesting that morphological characters often reflect a pattern similar to that seen with molecules. Greater numbers of molecular characters can overwhelm morphological characters (Hillis, 1987), depending on relative numbers of informative characters, but such a threat is not as dangerous as it seems (Wenzel and Siddall, 1999). Even small numbers of morphological characters can contribute significantly to the results of a combined analysis (Barrett et al., 1991); as Donoghue and Sanderson (1992) pointed out, the sheer numbers of characters are not as important as character interaction and distribution of homoplasy. Gatesy and Arctander (2000) found that

morphological characters provided over half of the partitioned branch support in their simultaneous analysis of five datasets.

Goodman et al. (1987, p. 147) stated that, “A more basic problem with morphological characters as indicators of genealogical relationships is that there is no direct correspondence between the characters and heritable information encoded in genomic DNA.” Clearly, this statement is questionable when so broadly framed, as there are many simple characters (flower color, for example) whose genetic basis is known to be straightforward (see Gottlieb (1984) for examples in plants). It is particularly in more complex (= polygenic) structures that the directly observable correspondence with underlying genetics may diminish (Roth, 1994). The issue of complexity has been considered in the systematic context of homology hypotheses (Riedl, 1978; McShea, 1991; Donoghue, 1992; Donoghue and Sanderson, 1994; Janies and DeSalle, 1999; Bang et al., 2000)—in particular the idea that more complex characters are likely to exhibit less homoplasy, which might argue for their use when compared to simple sequence characters.

Most of these discussions of molecules and morphology predate the era in which sequencing of whole genomes was a possibility. The fact that most combined analyses comprise very small portions of the genome means that the independence of these characters is a safe assumption, such that the decision about whether to analyze both morphological and molecular data is one of efficacy rather than redundancy of information. Morphology represents one end of a range of phenotypic expressions of the genome, which also includes nucleotide class (purine vs pyrimidine), amino acid, amino acid class, and higher-level structures in proteins and RNA. Given that they are all encoded by the genome, is there any benefit in coding potentially redundant, nonindependent phenotypic characters in addition to genomic characters in a phylogenetic analysis? We argue that there is, precisely because homology can be expressed at multiple different levels, from nucleotides to genes, gene functions, gene networks, embryonic origins, and morphological structures (Dickinson, 1995; Abouheif, 1999). We further argue that such a coding scheme is in fact a type of total-evidence approach (Kluge, 1989), because it takes advantage of homology hypotheses (characters) at all levels.

To the extent that characters are genomically determined, their state transformations will be marked by changes in the underlying gene sequence. This means that the genome should represent a record of all phenotypic changes exhibited by the organism and changes in the DNA sequence that may not cause any detectable change in the phenotype. Therefore, the question is really not whether the genome contains all of the information reflected in the phenotype, but whether it can be

recovered, which is fundamentally an issue of homology and transformation. Relative rates of change in the genome and phenotype are important here. Because not all changes that occur in the genome are reflected in the phenotype, the rate of change in the genome is expected to be higher, such that saturation of changes (i.e., multiple changes for characters along individual branches) may be exhibited in the genome relative to the phenotype. Saturation is most often discussed with reference to particular base positions, such as third codon base positions in coding sequences (e.g., Hillis, 1991), but the same effect can be observed with reference to any character if it changes often enough. The concern is that “noise” will result from characters that have changed so rapidly that it is difficult to reconstruct their transformations (though noise is relative; Wenzel and Siddall, 1999). While model-based phylogeny reconstruction methods use specific transformational assumptions to approach this problem, the ability to reconstruct the transformation of characters in a parsimony framework depends in large part on sufficient taxon sampling, because real specimens represent real character combinations that help to exclude some of the universe of possible character transformations between divergent taxa. If all intermediates were available and taxon sampling were complete, reconstruction of transformations would be more straightforward. When taxon sampling is sparse, transformation reconstruction becomes more difficult and is facilitated by the addition of characters that change more slowly and less ambiguously partition the taxa (e.g., Davis et al., 1998). Slowly changing or “conservative” characters are often prized for their quality as phylogenetic markers (e.g., Lloyd and Calder, 1991; Lanyon, 1988). Felsenstein’s (1981) proposed weighting of characters based on their improbability is a codification of what has long been practiced intuitively.

The absence of a lockstep correspondence between a complex phenotype and underlying genetics can be used to advantage in systematic studies, since phenotypic characters may retain evidence for homology when the underlying genotypic characters do not (de Beer, 1971; Meyer, 1999; Wray, 1999; but see Doyle, 1996, p. 59). As Meyer (1999, p. 144) noted, “nonhomologous genes, gene networks and developmental mechanisms can make structures that are typically considered to be homologues.” This may occur, for instance, in a biosynthetic pathway when one protein is substituted for another protein that is coded by a different gene. Kjer (1995) noted the more highly conserved nature of ribosomal RNA secondary structure relative to nucleotide sequence. This retention of homology is particularly important when distantly related taxa are sampled (due to extinction or undersampling). As one moves up the hierarchy of these levels (from nucleotide to morphological structure), one may generally expect the higher-level characters to evolve

more slowly than the lower-level characters, because each dependent higher-level character and/or character state may include multiple dependent lower-level characters and/or character states, as noted above. Hence, synapomorphies obscured by multiple changes at the lower-level character(s) may be retained by the higher-level character(s). In such cases, phylogenetic signal may be more easily recoverable from the phenotypic than from the genotypic characters (Lanyon, 1988; Naylor and Brown, 1998). Phenotypic characters can serve the function of “guiding” the genotypic characters as transformations are reconstructed by providing reinforcement for key genomic synapomorphies. Hence, like increasing taxon sampling (Hendy and Penny, 1989; Hillis, 1996; Phillippe et al., 1996; Graybeal, 1998; Zwickl and Hillis, 2002), addition of phenotypic characters to analyses including genotypic characters on which they are based can help to clarify character state transformations. Although this approach may seem unusual to phylogeneticists, developmental biologists have recognized the need to bridge the gap between sequence and morphology and have begun to comment on the phylogenetic utility of simultaneous analysis of characters derived from different developmental levels (Janies and DeSalle, 1999; Bang et al., 2000).

The level of genomic transmission may also include some epigenetic changes, although the importance of such changes in marking phylogenetic pattern remains unclear. Various meanings exist for the term “epigenetic” (Bird, 1998); Waddington (1939) defined the term to mean the complex interactions that comprise the development of a multicellular organism. Some of these interactions lead to changes in the genome that may be heritable, but that are not due to a change in base sequence (Jablonka and Lamb, 1998; Cubas et al., 1999; Wolffe and Matzke, 1999)—they are due instead to methylation or changes in chromatin structure (Wolffe and Matzke, 1999). *Epigenetics* is now often used to refer to the study of such heritable but nonsequence-based changes (Bird, 1998). When such changes are heritable and fixed in taxa, they are eligible to be used as characters. To the extent that they are not fixed or heritable, they may appear as polymorphisms, which are problematic for phylogenetic analysis (Nixon and Davis, 1991; Mabee and Humphries, 1993).

Analysis of phenotype and genotype

The example provided in Fig. 2 compares the information contributed by phenotypic and genotypic levels—namely amino acids and their underlying sequences. In this example, the third position of a codon varies among six taxa. Assuming that the states in Taxon 1 are plesiomorphic, cladistic analysis of just the nucleotide sequences (under Fitch parsimony) would

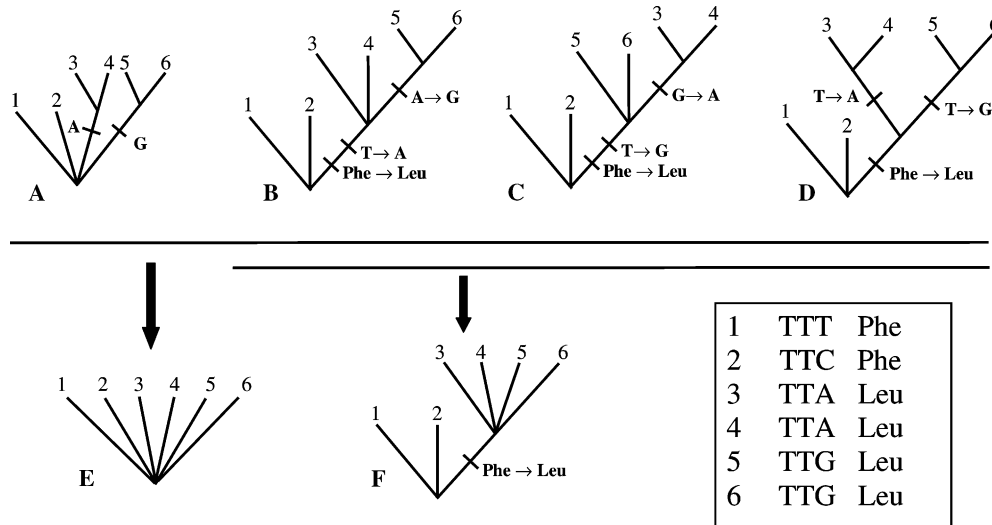


Fig. 2. Use of genomic and phenotypic data. A simple data matrix with four sequences is translated into amino acids. (A–D) Individual trees resulting from cladistic analysis of the data. Trees A–D are obtained when only nucleotides are analyzed; trees B–D result from analysis of nucleotides plus the amino acid character. (E) Strict consensus of trees A–D. (F) Strict consensus of trees B–D.

result in the trees shown in Figs. 2A–D and the strict consensus polytomy in Fig. 2E. However, if the amino acid character is added to the analysis, the trees in Figs. 2B–D and consensus tree in Fig. 2F result. The amino acid character has a clear synapomorphy, whereas the sequence data set does not. There is no incongruence among the trees, but addition of the amino acid character allows the exclusion of some possible trees from the results, thus clarifying the transformation.

Agosti et al. (1996) introduced the simultaneous use of nucleotide and amino acid characters derived from the same sequence in phylogenetic analysis of protein-coding genes. By using nucleotide and amino acid characters, information may be incorporated from both methods of coding the sequence data. Agosti et al. (1996, p. 67) recognized three possible outcomes when nucleotide and amino acid characters are coded:

First, each amino acid and the information in its associated triplet may be entirely congruent and equally informative with respect to each other. Second, one of the sources of information will show no informativeness while the other will. Finally, the two sources of information—a nucleic acid triplet and its amino acid—may be informative but incongruent.

There is also a fourth possibility: both sources of information will be congruent, but show different char-

acter state distributions. Examples of these four possibilities are illustrated in Table 1. The third and fourth possibilities, in which the nucleotide and amino acid characters are either incongruent or nonredundant, may be caused by silent substitutions in one or more of the nucleotide characters, convergence of the amino acid character (Simmons, 2000), or an artificial character state caused by the use of composite coding for the amino acid character (Simmons and Freudenstein, 2002). For the first of these three possibilities, incorporation of the amino acid character is advantageous because direct evidence for the synapomorphy is retained with the amino acid character, but lost with the nucleotide characters because of subsequent silent substitutions. The second possibility is a potential problem whenever multiple distinct character states are grouped into a single character state (e.g., coding bases when only considering transversions). The third possibility is a potential problem whenever separate nucleotide characters are grouped together into a single character (e.g., morphological characters). These two problems have been shown to cause artificial resolution with empirical data (Simmons et al., 2002). Unfortunately, there is no way to distinguish between these three possibilities without reference to a phylogenetic tree. Therefore, it is

Table 1
Nucleotide character states for four codons with their corresponding amino acids in parentheses

	Codon 1	Codon 2	Codon 3	Codon 4
Taxon 1	AGT (S)	AGT (S)	AGT (S)	AGT (S)
Taxon 2	AGT (S)	AGT (S)	TCC (S)	AGT (S)
Taxon 3	AGT (S)	AGT (S)	GAT (D)	AGC (S)
Taxon 4	AGA (R)	AGC (S)	GAC (D)	AGA (R)
Taxon 5	AGA (R)	AGC (S)	GAC (D)	AGA (R)

impossible a priori to determine when the use of higher-level characters is advantageous or disadvantageous.

Agosti et al. (1996) and Birstein and DeSalle (1998) considered their method to be an objective way to up-weight congruent, redundant characters such that slowly evolving, nonsynonymous substitutions may be given twice the weight of more rapidly evolving synonymous substitutions, the former considered to be more reliable than the latter. Other workers have also considered this method of up-weighting to be advantageous (Buck et al., 2000; Flores-Villela et al., 2000). However, silent substitutions may actually be better than replacement substitutions for phylogenetic reconstruction (Simmons et al., 2002). Furthermore, the assumption that slowly evolving nucleotide characters are more reliable than faster evolving nucleotide characters has been thoroughly refuted (Yang, 1996; Yoder et al., 1996; Lewis et al., 1997; Björklund, 1999; Källersjö et al., 1999; Wenzel and Siddall, 1999; Baker et al., 2001). In general we believe that all characters should be equally weighted in tree searches because each character represents an independent hypothesis of relationship. We advocate use of this selective weighting method as modified below because of its potential for improving the reconstruction of transformations. This procedure is weighting in the sense that certain character state changes will be counted twice—however, it differs from what is normally termed weighting in that whole characters are not duplicated (see below) and because the weights are based on the data.

The Agosti et al. (1996) approach to including dependent characters in phylogenetic analysis can be extended to multiple different levels. For example, Flores-Villela et al. (2000) used nucleotide, base, and amino acid characters for each gene together in their phylogenetic analysis. We propose extending this method to the coding of *all* phenotypic levels in addition to genotype, followed by simultaneous analysis, to take advantage of the homologies retained at different phenotypic levels. Other phenotypic levels for sequence data that might be encoded include class of amino acid, secondary, tertiary, and quaternary (for oligomeric proteins) structures, the function of the protein in the biosynthetic pathway(s), and the morphological structure(s) to which the protein contributes. Primary homology statements may be made at every one of these levels. Note that we do not advocate the use of phenotypic characters to the exclusion of sequence characters when the latter are available; this approach would incur the same problems discussed by Simmons and Freudenstein (2002) that are encountered when using only composite characters as opposed to their reductive counterparts (Wilkinson, 1995).

Admittedly, inclusion of both phenotypic and genotypic characters in a single analysis will necessarily lead to redundancy. By redundancy we mean the repetition of information derived from nonindependent characters.

Redundancy of character information is not limited to the case where genotype and phenotype are utilized simultaneously, however, since it can also occur among genotypic characters where lack of independence occurs. Redundancy among genotypic characters is slightly different from that between genotype and phenotype. In the latter, redundancy is due to use of the same information through different phenotypic filters, whereas in the former it is due to duplication of the information in another physical location of the genome. Examples of redundancy within the genome include compensatory changes in stem regions of molecules with secondary structure (e.g., rRNA genes, Wheeler and Honeycutt, 1988; Type II introns, Michel et al., 1989) and repeats that have undergone concerted evolution (e.g., rDNA, Arnheim, 1983; ubiquitin, Sharp and Li, 1987; histone genes, Baldo et al., 1999). If the entire genome is utilized in an analysis, stem regions and repeats will all be included, hence multiplying this information. The problem of potential nonindependence and resulting duplication of information in stem regions has long been recognized, leading some to advise down-weighting these regions in analyses (Wheeler and Honeycutt, 1988; Dixon and Hillis, 1993; Springer et al., 1995) to avoid redundancy.

How can we alleviate the problem of redundancy (and consequent weighting) encountered in coding dependent characters? If this problem is not addressed, a replacement nucleotide substitution that affected the character coding of all nine levels cited above would be given nine times the weight of a silent transition. To address this weighting problem in our method, we introduce a novel approach that allows for the nonredundant coding of dependent characters. This approach is implemented by searching for identical character state distributions among dependent characters. The higher-level characters that have character state distributions identical with those of any of the lower-level character(s) on which they are demonstrably dependent are then deactivated. An example for a single codon, with four levels at which primary homology statements are considered, is presented in Table 2. The first and third codon positions are variable, but only the third position is parsimony informative. The character state distributions of third codon position base type and nucleotide are nonredundant. Similarly, the character state distribution of amino acid is nonredundant with those of any of the three nucleotide or base type characters. However, the state distribution of amino acid class is redundant with that of the amino acid character. Therefore, the amino acid class character is deactivated because it is redundant with the lower-level amino acid character on which it is dependent. All of the other characters would be equally weighted in the phylogenetic analysis.

Although we suggest that nonredundant coding of dependent characters be applied to all of the levels at which primary homology statements are appropriately

Table 2

Nonredundant coding of dependent characters for a single codon with four levels at which primary homology statements are considered

	Codon	Base (1st position)	Base (3rd position)	Amino acid	Amino acid class
Taxon 1	TTT	pyrimidine	pyrimidine	phenylalanine	aromatic, hydrophobic
Taxon 2	TTC	pyrimidine	pyrimidine	phenylalanine	aromatic, hydrophobic
Taxon 3	TTA	pyrimidine	purine	leucine	aliphatic, hydrophobic
Taxon 4	TTA	pyrimidine	purine	leucine	aliphatic, hydrophobic
Taxon 5	CTT	pyrimidine	pyrimidine	leucine	aliphatic, hydrophobic

made, we realize that this method may be difficult to apply at higher levels (above the level of class of amino acid). This is because multiple codons are involved in specifying the higher-level characters, and it will often be difficult to establish dependency between any given higher-level character and the multiple lower-level characters upon which it is based. For example, suppose it is known that 10 separate genes control a single morphological character, which is parsimony informative for the group being studied. Although the 10 genes that control the morphological character are known, which of the 10 genes or which region of any 1 gene is responsible for the variation recorded in the morphological character is not known. The character state distribution for the morphological character will almost certainly be redundant with the character state distributions of many of the lower-level characters upon which it is based (nucleotide characters, base characters, amino acid characters, etc. for the 10 genes). Given this limited knowledge and a loose interpretation of our nonredundant coding, the morphological character would be deactivated because of its identical character state distribution relative to that of some of the lower-level characters upon which it is based. However, because it is unclear whether the character state distribution for the morphological character is dependent on, or merely coincident with, any of those lower-level characters given this limited knowledge, we suggest that the morphological character should not be deactivated. Using this approach, a higher-level character is deactivated only when its character state distribution is *demonstrably* redundant with at least one of the lower-level characters upon which it is based.

An effect of making primary homology statements at multiple different levels is illustrated by an example of two genes of different lengths but with the same number of parsimony informative nucleotide characters. The higher-level characters can contribute very different numbers of parsimony informative characters to the phylogenetic analysis. This is because the faster-evolving gene is expected to have less redundancy among the multiple different levels at which primary homology statements are made than the slower-evolving gene. That is, the more multiple hits, the more phylogenetic information is potentially recoverable from a given gene when higher-level characters are incorporated. Just as this rule applies for phylogenetic analyses of more taxa rather than fewer taxa (such that globally homoplasious characters are locally

informative; Källersjö et al., 1999; Wenzel and Siddall, 1999), it applies to phylogenetic analyses in which primary homology statements are made at more levels.

We note a similarity in use of such levels simultaneously in cladistic analysis with the process of reciprocal illumination (e.g., Hennig, 1966; Pogue and Mickevich, 1990). In both processes, we compare composite or phenotypic characters with their corresponding reductive characters or components, in both cases relying on the possibility of the underlying characters being less conservative than their higher-level phenotypes. Such an approach is intrinsic to reciprocal illumination, where primary homology statements are reevaluated whenever characters are optimized as homoplasious on the most parsimonious tree(s) (de Pinna, 1991). This is done by reexamining, at a “lower” (or “finer”) level, the structures that were initially coded as the same character state. For example, the basis for a gross morphological character may be tested by examining anatomy or gene expression (Janies and DeSalle, 1999; but see Wray, 1999). If upon examination of the structures at a lower level they are deemed sufficiently different, the character is recoded such that the structures are no longer coded as homologous. Note, however, that this procedure constitutes selective implementation of the expectation of homology between higher- and lower-level characters. Whereas we emphasize here that higher-level characters can retain similarity when their component lower-level characters do not, in reciprocal illumination the higher-level characters are hypothesized to be homologous *only if* their component lower level characters are also homologous.

A problem that can occur with the non-redundant-coding method for higher-level characters that are based on two or more lower-level characters is that the higher-level character can be partially redundant with two or more lower-level, multistate characters upon which it is based. For example, in Fig. 3 the amino acid character for codon 1 is partially redundant with the nucleotide character for the first codon position. The amino acid character is demonstrably redundant with the nucleotide character for the first codon position when only taxa 1–7 are considered. However, because of the substitution at the third codon position in taxon 8, when taxa 1–8 are considered, the amino acid character is partially redundant with the nucleotide characters for the first and third codon positions, but is completely redundant with

	codon 1			codon 2			codon 3		
Taxon 1	TAT	(Y)	[?]	AGA	(R)	[R]	GGT	(G)	[?]
Taxon 2	TAT	(Y)	[?]	AGT	(S)	[?]	GGT	(G)	[?]
Taxon 3	TAT	(Y)	[?]	AGT	(S)	[?]	AGC	(S)	[S]
Taxon 4	CAT	(H)	[?]	AGT	(S)	[?]	AGT	(S)	[S]
Taxon 5	CAT	(H)	[?]	AGG	(R)	[R]	AGT	(S)	[S]
Taxon 6	CAT	(H)	[?]	AGA	(R)	[R]	AGT	(S)	[R]
Taxon 7	AAT	(N)	[N]	AGG	(R)	[R]	AGG	(R)	[R]
Taxon 8	AAG	(K)	[?]	AGA	(R)	[R]	AGA	(R)	[R]

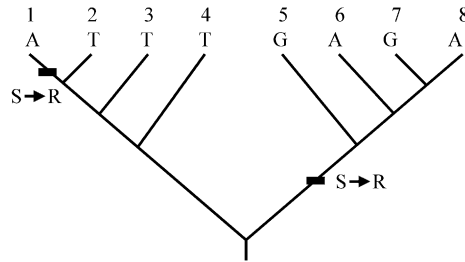


Fig. 3. Examples where the amino acid characters are partially redundant with the nucleotide characters. Amino acid translations are shown in parentheses, with the modified non-redundant-coding method (see text) shown in brackets. For codon 1, the amino acid character is partially redundant with the first and third codon positions, but is completely redundant with neither. For codon 2, the amino acid character is redundant with the third codon position in one part of the tree but not in another. The character states for the third base position of the second codon are listed below the taxon numbers on the tree. The character state changes for the amino acid character for codon 2 are mapped onto the tree. Inserting a “?” for the amino acid state of taxa 2–4 eliminates the partial redundancy, but renders the entire amino acid character uninformative. For codon 3, the amino acid character is partially redundant with the first base position; inserting a “?” for taxa 1 and 2 eliminates the partial redundancy, but does not render the amino acid character uninformative. The remaining amino acid character state distribution for codon 3 provides a signal different from that of any of the lower-level nucleotide characters.

neither. Although it would be appropriate to deactivate the amino acid character in such a case, the non-redundant-coding method would fail to do so. To address this problem, the non-redundant-coding method may be modified such that individual character states are rescored as missing data for the higher-level character that had the same distribution as any particular character state for any of the lower-level characters on which it was based. For codon 1, this would result in an uninformative amino acid character in which Taxon 7 would be scored as “N” and all other taxa would be scored as “?” for the amino acid character.

Cases of partial redundancy may also arise when multiple hits accrue in one portion of the tree but not in another. For example, in Fig. 3, the amino acid character for codon 2 is redundant with the nucleotide character for the third codon position on the left side of the tree but not on the right side of the tree. Because of the silent substitutions in the clade consisting of terminals 5–8, the third codon position nucleotide changes are ambiguously optimized on the right side of the tree. In contrast, the amino acid change is unambiguously optimized as a synapomorphy for the clade. The subsequent silent substitutions have obscured the replacement substitution at the base of the clade for the nucleotide character, but not the amino acid character. The potential modification of the nonredundant coding method mentioned above would eliminate this problem, but in doing so would render the amino acid character

uninformative. However, the modification need not always render the partially redundant higher-level character uninformative. Codon 3 of Fig. 3 shows an example where the amino acid character is partially redundant with the nucleotide character for the first codon position, but the modified non-redundant-coding method would only deactivate the amino acid character states for taxa 1 and 2, leaving the character states that provide a transition from R to S.

For efficient data matrix construction, nonredundant coding of dependent characters could be implemented in computer programs, such as MacClade (Maddison and Maddison, 1992) and WinClada (Nixon, 1999). Dependent characters would be given unique tags and the program would check for identical character state distributions within each set of tagged characters. The redundant, higher-level characters would then be deactivated. This would operate in a manner similar to that of the “compress character” command in MacClade and the “pack” command in Nona (Goloboff, 1993), which are used to speed tree searches.

Conclusion

Although complete genome sequences can provide us with all of the genetic characters, we can never have all of the taxa that are relevant to a given phylogenetic analysis (due to undersampling and extinction of

ancestral and “intermediate” terminals). Therefore, we can lose synapomorphies with nucleotide characters (because of multiple hits), but retain them with phenotypic characters (i.e., patterns that may not be observable at the nucleotide level may be observable at a given phenotypic level). The central importance of the method outlined here is that it attempts to maximize the amount of homology information that we can extract from the data at hand. This is useful because it improves reconstruction of character transformations on the tree. In this way, its benefit is similar to increasing taxon sampling. Beyond sequences and their derivative phenotypic levels, nongenomically based characters should not be overlooked, since they provide an additional source of historical markers that can only serve to further strengthen our phylogenetic hypotheses.

Acknowledgments

We thank Joel Cracraft, Rob DeSalle, Arnold Kluge, Daniel Potter, Martin Ramirez, Ward Wheeler, an anonymous reviewer, and the OSU Phylogenetics Discussion Group for discussion and comments on the manuscript.

References

- Abouheif, E., 1999. Establishing homology criteria for regulatory gene networks: prospects and challenges. In: Bock, G.R., Cardew, G. (Eds.), *Homology*. Wiley, Chichester, pp. 207–225.
- Agosti, D., Jacobs, D., DeSalle, R., 1996. On combining protein sequences and nucleic acid sequences in phylogenetic analysis: the homeobox protein case. *Cladistics* 12, 65–82.
- Arnheim, N., 1983. Concerted evolution of multigene families. In: Nei, M., Koehn, R.K. (Eds.), *Evolution of Genes and Proteins*. Sinauer, Sunderland, MA, pp. 38–61.
- Allard, M.W., Farris, J.S., Carpenter, J.M., 1999. Congruence among mammalian mitochondrial genes. *Cladistics* 15, 75–84.
- Baker, R.H., Wilkinson, G.S., DeSalle, R., 2001. Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsideae). *Syst. Biol.* 50, 87–105.
- Balaban, E., 1988. Cultural and genetic variation in swamp sparrows (*Melospiza georgiana*). I. Song variation, genetic variation and their relationship. *Behaviour* 105, 250–291.
- Bang, R., DeSalle, R., Wheeler, W., 2000. Transformationalism, taxism, and developmental biology in systematics. *Syst. Biol.* 49, 19–27.
- Baldo, A.M., Les, D.H., Strausbaugh, L.D., 1999. Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of *Sophophora*. *Mol. Biol. Evol.* 16, 1511–1520.
- Barrett, M., Donoghue, M.J., Sober, E., 1991. Against consensus. *Syst. Zool.* 40, 486–493.
- Bird, A., 1998. Comments on epigenetic inheritance in evolution. *J. Evol. Biol.* 11, 185–188.
- Birstein, V.J., DeSalle, R., 1998. Molecular phylogeny of Acipenserinae. *Mol. Phylogenet. Evol.* 9, 141–155.
- Björklund, M., 1999. Are third positions really that bad? A test using vertebrate cytochrome *b*. *Cladistics* 15, 191–197.
- Bollyky, P.L., Holmes, E.C., 1999. Reconstructing the complex evolutionary history of Hepatitis B virus. *J. Mol. Evol.* 49, 130–141.
- Brodo, I.M., Sharnoff, S.D., Sharnoff, S., 2001. *Lichens of North America*. Yale University Press, New Haven.
- Brooks, D.R., 1981. Hennig’s parasitological method: a proposed solution. *Syst. Zool.* 30, 229–249.
- Brown, D.L., Knox, J.D., Paulin–Levasseur, M., 1992. The lymphocyte centrosome. In: Kalninis, V.I. (Ed.), *The Centrosome*. Academic Press, New York, pp. 261–286.
- Brown, J.R., 1996. Preparing for the flood: evolutionary biology in the age of genomics. *Trends Ecol. Evol.* 11, 510–513.
- Buck, W.R., Goffinet, B., Shaw, A.J., 2000. Testing morphological concepts of orders of pleurocarpous mosses (Bryophyta) using phylogenetic reconstructions based on *TRNL-TRNF* and *RPS4* sequences. *Mol. Phylogenet. Evol.* 16, 180–198.
- Callaini, G., Ripabelli, M.G., Dallai, R., 1999. Centrosome inheritance in insects: fertilization and parthenogenesis. *Biol. Cell* 91, 355–366.
- Chapela, I.H., Rehner, S., Schultz, T.R., Mueller, U., 1994. Evolutionary history of the symbiosis between fungus-growing ants and their fungi. *Science* 266, 1691–1694.
- Clausen, J., Keck, D.D., Hiesey, W.M., 1940. *Experimental Studies on the Nature of Species. I. Effect of Varied Environments on Western North American Plants*. Carnegie Institution of Washington Publ., p. 520.
- Cleveland, L.R., 1926. Symbiosis among animals with special reference to termites and their intestinal flagellates. *Quart. Rev. Biol.* 1, 51–60.
- Collias, N.E., Collias, E.C., 1984. *Nest Building and Bird Behavior*. Princeton University Press, Princeton, NJ.
- Collinge, J., Sidle, K.C.L., Meads, J., Ironside, J., Hill, A.F., 1996. Molecular analysis of prion strain variation and the aetiology of ‘new variant’ CJD. *Nature (Lond.)* 383, 685–690.
- Corliss, J.O., 1979. *The Ciliated Protozoa: Characterization, Classification and Guide to the Literature*. Pergamon Press, Oxford.
- Cox, B.S., 1965. A cytoplasmic suppressor of super-suppressor in yeast. *Heredity* 24, 505–521.
- Cubas, P., Vincent, C., Coen, E., 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature (Lond.)* 401, 157–161.
- Davis, I.J., Simmons, M.P., Stevenson, D.W., Wendel, J.F., 1998. Data decisiveness, data quality, and incongruence in phylogenetic analysis: an example from the monocotyledons using mitochondrial *atp A* sequences. *Syst. Biol.* 47, 282–310.
- Dawkins, R., 1976. *The Selfish Gene*. Oxford University Press, Oxford.
- de Beer, G., 1971. *Homology, an Unsolved Problem*. Oxford University Press, London.
- de Pinna, M.C.C., 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7, 367–394.
- Dickinson, W.J., 1995. Molecules and morphology: where’s the homology? *Trends Genet.* 11, 119–121.
- Dixon, M.T., Hillis, D.M., 1993. Ribosomal RNA secondary structure–compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* 10, 256–267.
- Donoghue, M.J., 1992. Homology. In: Keller, E.F., Lloyd, E.A. (Eds.), *Keywords in Evolutionary Biology*. Harvard University Press, Cambridge, MA, pp. 170–179.
- Donoghue, M.J., Sanderson, M.J., 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. In: Soltis, P.S., Soltis, D.E., Doyle, J.J. (Eds.), *Molecular Systematics of Plants*. Chapman and Hall, New York, pp. 340–368.
- Donoghue, M.J., Sanderson, M.J., 1994. Complexity and homology in plants. In: Hall, B.K. (Ed.), *Homology: The Hierarchical Basis of Comparative Biology*. Academic Press, San Diego, pp. 393–421.
- Doyle, J.J., 1991. Gene trees and species trees: molecular systematics as one-character taxonomy. *Sys. Bot.* 17, 144–163.
- Doyle, J.J., 1996. Homoplasmy connections and disconnections: genes and species, molecules and morphology. In: Sanderson, M.J., Hufford, L. (Eds.), *Homoplasmy: The Recurrence of Similarity in Evolution*. Academic Press, San Diego, pp. 37–66.

- Dressler, R.L., 1990. The major clades of the Orchidaceae–Epidendroideae. *Lindleyana* 5, 117–125.
- Eichler, W., 1941. Wirtsspezifität und stammesgeschichtliche Gleichläufigkeit Fahrenholzschke (Regel) bei Parasiten im allgemeinen und bei Mallophagen im besonderen. *Zool. Anz.* 132, 254–262.
- Felsenstein, J., 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* 16, 183–196.
- Fleischer, R.C., Rothstein, S.I., 1988. Known secondary contact and rapid gene flow among subspecies and dialects in the brown-headed cowbird. *Evolution* 42, 1146–1158.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., Venter, J.C., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.
- Flores-Villela, O., Kjer, K.M., Benabib, M., Sites, J.W., 2000. Multiple data sets, congruence, and hypothesis testing for the phylogeny of basal groups of the lizard genus *Sceloporus* (Squamata, Phrynosomatidae). *Syst. Biol.* 49, 713–739.
- Fraser, C.M., Eisen, J.A., Salzberg, S.L., 2000. Microbial genome sequencing. *Nature (Lond.)* 406, 799–803.
- Freudenstein, J.V., 1999. Relationships and character transformation in Pyroloideae (Ericaceae) based on ITS sequences, morphology, and development. *Syst. Bot.* 24, 398–408.
- Gatesy, J., Arctander, P., 2000. Hidden morphological support for phylogenetic placement of *Pseudoryx nghetinhensis* with the bovine Bovids: a combined analysis of gross anatomical evidence and DNA sequences from five genes. *Syst. Biol.* 49, 515–538.
- Gatesy, J., DeSalle, R., Wheeler, W., 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Mol. Phylogenet. Evol.* 2, 152–157.
- Gaunt, S.L.L., Baptista, L.F., Sanchez, J.E., Hernandez, D., 1994. Song learning as evidenced from song sharing in two hummingbird species (*Colibri coruscans* and *C. thalassinus*). *Auk* 111, 87–103.
- Givnish, T.J., Sytsma, K.J., 1997a. Consistency, characters, and the likelihood of correct phylogenetic inference. *Mol. Phylogenet. Evol.* 7, 320–333.
- Givnish, T.J., Sytsma, K.J., 1997b. Homoplasy in molecular vs. morphological data: the likelihood of correct phylogenetic inference. In: Givnish, T.J., Sytsma, K.J. (Eds.), *Molecular Evolution and Adaptive Radiation*. Cambridge University Press, Cambridge, pp. 55–101.
- Goloboff, P.A., 1993. “*Nona*, version 1.6 (computer software and manual)”. Distributed by the author.
- Goodman, M., Miyamoto, M.M., Czelusniak, J., 1987. Pattern and process in vertebrate phylogeny revealed by coevolution of molecules and morphologies. In: Patterson, C. (Ed.), *Molecules and Morphology in Evolution: Conflict or Compromise?* Cambridge University Press, Cambridge, pp. 141–176.
- Gottlieb, L.D., 1984. Genetics and morphological evolution in plants. *Am. Nat.* 123, 681–709.
- Gottlieb, L.D., 1988. Towards molecular genetics in *Clarkia*: gene duplications and molecular characterization of PGI genes. *Ann. Mo. Bot. Gard.* 75, 1169–1179.
- Govan, V.A., Leat, N., Allsopp, M., Davison, S., 2000. Analysis of the complete genome sequence of acute bee paralysis virus shows that it belongs to the novel group of insect-infecting RNA viruses. *Virology* 277, 457–463.
- Graur, D., 1993. Molecular phylogeny and the higher classification of eutherian mammals. *Trends Ecol. Evol.* 8, 141–147.
- Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47, 9–17.
- Hedges, S.B., Maxson, L.R., 1996. Re: Molecules and morphology in amniote phylogeny. *Mol. Phylogenet. Evol.* 6, 312–314.
- Hedges, S.B., Sibley, C.G., 1994. Molecules vs. morphology in avian evolution: the case of the “pelecaniform” birds. *Proc. Natl. Acad. Sci. USA* 91, 9861–9865.
- Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 297–309.
- Hennig, W., 1966. *Phylogenetic Systematics*. University of Illinois Press, Urbana.
- Hillis, D.M., 1987. Molecular versus morphological approaches to systematics. *Ann. Rev. Ecol. Syst.* 18, 23–42.
- Hillis, D.M., 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. In: Miyamoto, M.M., Cracraft, J. (Eds.), *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, New York, pp. 278–294.
- Hillis, D.M., 1996. Inferring complex phylogenies. *Science* 383, 130–131.
- Hofstetter, H., Monstein, H.-J., Weissmann, C., 1974. *Biochim. Biophys. Acta* 374, 238–244.
- Huynen, M.A., Bork, P., 1998. Measuring genome evolution. *Proc. Natl. Acad. Sci. USA* 95, 5849–5856.
- Jablonska, E., Lamb, M.J., 1998. Epigenetic inheritance in evolution. *J. Evol. Biol.* 11, 159–183.
- Janies, D., DeSalle, R., 1999. Development, evolution, and corroboration. *Anat. Rec.* 257, 6–14.
- Johannsen, W., 1909. *Elemente der exakten Erblichkeitslehre*. Fischer, Jena.
- Källersjö, M., Albert, V.A., Farris, J.S., 1999. Homoplasy increases phylogenetic structure. *Cladistics* 15, 91–93.
- Kjer, K.M., 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Mol. Phylogenet. Evol.* 4, 314–330.
- Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38, 7–25.
- Kroodsma, D.E., Baker, M.C., Baptista, L.F., Petrino, L., 1985. Vocal “dialects” in Nuttall’s white-crowned sparrow. *Curr. Ornithol.* 2, 103–133.
- Lacroute, F., 1971. Non-Mendelian mutation allowing ureidosuccinic acid uptake in yeast. *J. Bacteriol.* 206, 519–522.
- Lanyon, S.M., 1988. The stochastic mode of molecular evolution: what consequences for systematic investigations? *Auk* 105, 563–573.
- Lewis, L.A., Mishler, B.D., Vilgalys, R., 1997. Phylogenetic relationships of the liverworts (Hepaticeae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Mol. Phylogenet. Evol.* 7, 377–393.
- Li, K., Xu, E.Y., Cecil, J.K., Turner, F.R., Megraw, T.L., Kaufman, T.C., 1998. *Drosophila* centrosomin protein is required for male meiosis and assembly of the flagellar axoneme. *J. Cell Biol.* 141, 455–467.
- Lindquist, S., Patino, M.M., Chernoff, Y.O., Kowal, A.S., Singer, M.A., Liebman, S.W., Lee, K.H., Blake, T., 1995. The role of Hsp104 in stress tolerance and [PSI⁺] propagation in *Saccharomyces cerevisiae*. *Cold Spring Harbor Symp. Quant. Biol.* 60, 451–460.
- Lindquist, S., Krobisch, S., Liming, L., Sondheimer, N., 2001. Investigating protein conformation-based inheritance and disease in yeast. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 169–176.
- Lindstrom, S.E., Hiromoto, Y., Nerome, R., Omoe, K., Sugita, S., Yamazaki, Y., Takahashi, T., Nerome, K., 1998. Phylogenetic analysis of the entire genome of Influenza A (H3N2) viruses from Japan: evidence for genetic reassortment of the six internal genes. *J. Virol.* 72, 8021–8031.

- Lloyd, D.G., Calder, V.L., 1991. Multi-residue gaps, a class of molecular characters with exceptional reliability for phylogenetic analyses. *J. Evol. Biol.* 4, 9–21.
- Louheed, S.C., Handford, P., 1992. Vocal dialects and the structure of geographic variation in morphological and allozymic characters in the rufous-collared sparrow, *Zonotrichia capensis*. *Evolution* 46, 1443–1456.
- Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Syst. Biol.* 49, 628–651.
- Mabee, P.M., Humphries, J., 1993. Coding polymorphic data: examples from allozymes and ontogeny. *Syst. Biol.* 42, 166–181.
- Maddison, W.P., Maddison, D.R., 1992. *MacClade: Analysis of Phylogeny and Character Evolution*. Sinauer, Sunderland, MA.
- Magasanik, B., 1992. Regulation of nitrogen utilization. In: Jones, E.W., Pringle, J.R., Broach, J.R. (Eds.), *The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Gene Expression, vol. 2*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 283–317.
- Makarova, K.S., Aravind, L., Galperin, M.Y., Grishin, N.V., Tatusov, R.L., Wolf, Y.I., Koonin, E.V., 1999. Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* 9, 608–628.
- Marler, P., Tamura, M., 1964. Culturally transmitted patterns of vocal behavior in sparrows. *Science* 146, 1483–1486.
- Marler, P., Peters, S., 1981. Sparrows learn adult song and more from memory. *Science* 213, 780–782.
- Martin, M.M., 1987. *Invertebrate-microbial Interactions: Ingested Fungal Enzymes in Arthropod Biology*. Comstock, Cornell University Press, Ithaca.
- McShea, D.W., 1991. Complexity and evolution—what everybody knows. *Biol. Philos.* 6, 303–324.
- Meyer, A., 1999. Homology and homoplasy: the retention of genetic programmes. In: Bock, G.R., Cardew, G. (Eds.), *Homology*. Wiley, Chichester, pp. 141–157.
- Michel, F., Umesono, K., Ozeki, H., 1989. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82, 5–30.
- Miller, J.S., Wenzel, J.W., 1995. Ecological characters and phylogenetic inference in the study of insects. *Ann. Rev. Entomol.* 40, 389–415.
- Miller, W., 2000. So many genomes, so little time. *Nat. Biotechnol.* 18, 148–149.
- Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C., Yuri, T., 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Syst. Biol.* 48, 138–152.
- Mundinger, P.C., 1979. Call learning in the Carduelinae: ethological and systematic considerations. *Syst. Zool.* 28, 270–283.
- Naylor, G.J.P., Brown, W.M., 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparison of sequences. *Syst. Biol.* 47, 61–76.
- Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., Fraser, C.M., 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature (Lond.)* 399, 323–329.
- Newman, S.A., Müller, G.B., 2000. Epigenetic mechanisms of character origination. *J. Exp. Zool.* 288, 304–317.
- Nicklas, R.B., 1971. Mitosis. In: Prescott, D.M., Goldstein, L., McConkey, J.E. (Eds.), *Advances in Cell Biology*, vol. 2. Appleton-Century-Crofts, New York, pp. 225–297.
- Nicklas, R.B., Koch, C.A., 1972. Chromosome micromanipulation. Part 4. Polarized motions within the spindle and models for mitosis. *Chromosoma* 39, 1–26.
- Nierman, W., Eisen, J.A., Fraser, C.M., 2000. Microbial genome sequencing 2000: new insights into physiology, evolution and expression analysis. *Res. Microbiol.* 151, 79–84.
- Nixon, K.C., 1999. *Winclada (BETA)*. Published by the author, Ithaca, NY.
- Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. *Cladistics* 12, 221–241.
- Nixon, K.C., Davis, J.I., 1991. Polymorphic taxa, missing values and cladistic analysis. *Cladistics* 7, 233–241.
- Palazzo, R.E., Vogel, J.M., Schnackenberg, B.J., Hull, D.R., Wu, X., 2000. Centrosome maturation. *Curr. Top. Dev. Biol.* 49, 449–470.
- Parchi, P. et al., 1996. Molecular basis of phenotypic variability in sporadic Creutzfeldt-Jakob disease. *Ann. Neurol.* 39, 669–680.
- Philippe, H., Adoutte, A., 1998. The molecular phylogeny of Eukaryota: solid facts and uncertainties. In: Coombs, G.H., Vickers, K., Sleigh, M.A., Warren, A. (Eds.), *Evolutionary Relationships Among Protozoa*. Chapman and Hall, London, pp. 25–56.
- Pogue, M.G., Mickevich, M.F., 1990. Character definitions and character state delineation: the bete noire of phylogenetic inference. *Cladistics* 6, 319–361.
- Prusiner, S.B., 1991. Molecular-biology of prion diseases. *Science* 252, 1515–1522.
- Prusiner, S.B., 1994. Biology and genetics of prion diseases. *Ann. Rev. Microbiol.* 48, 655–686.
- Rappaport, R., 1986. Establishment of the mechanism of cytokinesis in animal cells. *Int. Rev. Cytol.* 105, 245–281.
- Riedl, R., 1978. *Order in Living Organisms*. Wiley, New York.
- Roth, V.L., 1994. Within and between organisms: replicators, lineages and homologues. In: Hall, B.K. (Ed.), *Homology: The Hierarchical Basis of Comparative Biology*. Academic Press, San Diego, pp. 301–337.
- Schatten, G., 1994. The centrosome and its mode of inheritance: the reduction of the centrosome during gametogenesis and its restoration during fertilization. *Dev. Biol.* 165, 299–335.
- Serio, T.R., Lindquist, S.L., 2000. Protein-only inheritance in yeast: something to get [PSI⁺]-ched about. *Trends Cell Biol.* 10, 98–105.
- Sharp, P.M., Li, W.-H., 1987. Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. *J. Mol. Evol.* 25, 58–64.
- Sibley, C.G., Ahlquist, J.E., 1987. Avian phylogeny reconstructed from comparisons of the genetic material, DNA. In: Patterson, C. (Ed.), *Molecules and Morphology in Evolution: Conflict or Compromise?* Cambridge University Press, Cambridge, pp. 95–121.
- Simmons, M.P., 2000. A fundamental problem with amino-acid-sequence characters for phylogenetic analyses. *Cladistics* 16, 274–282.
- Simmons, M.P., Clevinger, C.C., Savolainen, V., Archer, R.H., Mathews, S., Doyle, J.J., 2001. Phylogeny of the Celastraceae inferred from phytochrome B gene sequence and morphology. *Am. J. Bot.* 88, 313–325.
- Simmons, M.P., Freudenstein, J.V., 2002. Artifacts of coding amino acids and other composite characters for phylogenetic analysis. *Cladistics* 18, 354–365.
- Simmons, M.P., Ochoterena, H., Freudenstein, J.V., 2002. Conflict between amino acid and nucleotide characters. *Cladistics* 18, 200–206.
- Sluder, G., 1992. Control of centrosome inheritance in echinoderm development. In: Kalnins, V.I. (Ed.), *The Centrosome*. Academic Press, New York, pp. 235–259.
- Smith, D.B., Basaras, M., Frost, S., Haydon, D., Cuceanu, N., Prescott, L., Kamenka, C., Millband, D., Sathar, M.A., Simmonds, P., 2000. Phylogenetic analysis of GBV-C/hepatitis G virus. *J. Gen. Virol.* 81, 769–780.

- Sondheimer, N., Lopez, N., Craig, E.A., Lindquist, S., 2001. The role of *Sis1* in the maintenance of the [RNQ(+)] prion. *EMBO J.* 20, 2435–2442.
- Springer, M.S., Hollar, L.J., Burk, A., 1995. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. *Mol. Biol. Evol.* 12, 1138–1150.
- Sytsma, K.J., Smith, J.F., Berry, P.E., 1991. Biogeography and evolution of morphology, breeding systems, flavonoids and chloroplast DNA in the four Old World species of *Fuchsia* (Onagraceae). *Syst. Bot.* 16, 257–269.
- Thornton, J.W., DeSalle, R., 2000. Gene family evolution and homology: genomics meets phylogenetics. *Ann. Rev. Genomics Human Genetics* 1, 41–73.
- Tram, U., Sullivan, W., 2000. Reciprocal inheritance of centrosomes in the parthenogenetic Hymenopteran *Nasonia vitripennis*. *Curr. Biol.* 10, 1413–1419.
- Tuite, M.F., 2000. Yeast prions and their prion-forming domain. *Cell* 100, 289–292.
- Tuite, M.F., Lindquist, S.L., 1996. Maintenance and inheritance of yeast prions. *Trends Genet.* 12, 467–471.
- Turlings, T.C.J., Wäckers, F.L., Vet, L.E.M., Lewis, W.J., Tuminson, J.H., 1993. Learning of host-finding cues by hymenopterous parasitoids. In: Papaj, D.R., Lewis, A.C. (Eds.), *Insect Learning*. Chapman and Hall, New York, pp. 51–78.
- Vrati, S., Giri, R.K., Razdan, A., Malik, P., 1999. Complete nucleotide sequence of an Indian strain of Japanese Encephalitis Virus: sequence comparison with other strains and phylogenetic analysis. *Am. J. Trop. Med. Hyg.* 61, 677–680.
- Vukmirovic, O.G., Tilghman, S.M., 2000. Exploring genome space. *Nature (Lond.)* 405, 820–822.
- Waddington, C.H., 1939. *Introduction to Modern Genetics*. Allen and Unwin, London.
- Weisenberg, R.C., 1972. Microtubule formation in vitro in solutions containing low calcium concentrations. *Science* 177, 1104–1105.
- Weissmann, C., 1994. Molecular biology of prion diseases. *Trends Cell Biol.* 4, 10–14.
- Wenzel, J.W., 1996. Learning, individual programs, and higher-level rules in construction behaviour of *Polistes*. In: Turillazzi, S., West-Eberhard, M.J. (Eds.), *Natural History and Evolution of Paper Wasps*. Oxford University Press, Oxford, pp. 58–74.
- Wenzel, J.W., Siddall, M.E., 1999. Noise. *Cladistics* 15, 51–64.
- Wheatley, D.N., 1982. *The Centriole: A Central Enigma of Cell Biology*. Elsevier Biomedical Press, New York.
- Wheeler, W.C., Honeycutt, R.L., 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Mol. Biol. Evol.* 5, 90–96.
- Whitfield, J.B., 2000. Phylogeny of microgastroid braconid wasps, and what it tells us about polydnavirus evolution. In: Austin, A.D., Dowton, M. (Eds.), *Hymenoptera: Evolution, Biodiversity and Biological Control*. CSIRO, Collingwood, Australia, pp. 97–105.
- Wickner, R.B., 1994. (URE3) as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* 264, 566–569.
- Wilkinson, M., 1995. A comparison of two methods of character construction. *Cladistics* 11, 297–308.
- Wolfe, A.P., Matzke, M.A., 1999. Epigenetics: regulation through repression. *Science* 286, 481–486.
- Wray, G.A., 1999. Evolutionary dissociations between homologous genes and homologous structures. In: Bock, G.R., Cardew, G. (Eds.), *Homology*. Wiley, Chichester, pp. 189–206.
- Wright, T.F., Wilkinson, G.S., 2001. Population genetic structure and vocal dialects in an Amazon parrot. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 609–616.
- Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* 42, 587–596.
- Yen, J.H., Barr, A.R., 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature (Lond.)* 232, 657–658.
- Yoder, A.D., Vilgalys, R., Ruvolo, M., 1996. Molecular evolutionary dynamics of cytochrome *b* in strepsirhine primates: the phylogenetic significance of third-position transversions. *Mol. Biol. Evol.* 13, 1339–1350.
- Zink, R., 1985. Genetical population structure and song dialects in birds. *Behav. Brain Sci.* 8, 118–119.
- Zink, R., Barrowclough, G.F., 1984. Allozymes and song dialects: a reassessment. *Evolution* 38, 444–448.
- Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51, 588–598.