# Incorporating Molecular Evolution into Phylogenetic Analysis, and a New Compilation of Conserved Polymerase Chain Reaction Primers for Animal Mitochondrial DNA

Chris Simon,[1,2] Thomas R. Buckley,[3] Francesco Frati,[4] James B. Stewart,[5,6] and Andrew T. Beckenbach[5]

[1]Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269; email: chris.simon@uconn.edu

[2]School of Biological Sciences, Victoria University of Wellington, Wellington 6014, New Zealand

[3]Landcare Research, Auckland 1142, New Zealand; email: BuckleyT@landcareresearch.co.nz

[4]Department of Evolutionary Biology, University of Siena, 53100 Siena, Italy; email: frati@unisi.it

[5]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada; email: jbs@alumni.sfu.ca, beckenba@sfu.ca

[6]Department of Laboratory Medicine, Division of Metabolic Diseases, Karolinska Institutet, Norvum 141 86, Stockholm, Sweden

Annu. Rev. Ecol. Evol. Syst. 2006. 37:545–79

First published online as a Review in Advance on August 16, 2006

The *Annual Review of Ecology, Evolution, and Systematics* is online at http://ecolsys.annualreviews.org

This article's doi: 10.1146/annurev.ecolsys.37.091305.110018



1543-592X/06/1201-0545$20.00

## Key Words

among-site rate variation, covarion-like evolution, molecular clocks, mtDNA genomes, nodal support, PCR primers

## Abstract

DNA data has been widely used in animal phylogenetic studies over the past 15 years. Here we review how these studies have used advances in knowledge of molecular evolutionary processes to create more realistic models of evolution, evaluate the information content of data, test phylogenetic hypotheses, attach time to phylogenies, and understand the relative usefulness of mitochondrial and nuclear genes. We also provide a new compilation of conserved polymerase chain reaction (PCR) primers for mitochondrial genes that complements our earlier compilation.

## INTRODUCTION

**mtDNA:** mitochondrial DNA

**Gene order rearrangement:** an evolutionary change in the location and/or direction of transcription of a gene with respect to other genes

**PCR:** polymerase chain reaction

**rRNA:** ribosomal RNA

The properties of the genes used and our ability to accommodate these properties have a much larger influence on the outcome of a molecular phylogenetic analysis than the particular method chosen to build a tree [although there are good reasons for preferring some phylogenetic methods over others (Holder & Lewis 2003; Swofford et al. 1996, 2001)]. It is from a careful study of data and their properties that empiricists can gain insight into the type of analyses needed (Simon 1991, Simon et al. 1994). Here we update our previous review of the evolution, weighting, and phylogenetic utility of mitochondrial genes and expand the focus from insects to all animals and from mitochondria to all DNA—although many of our examples still come from mitochondrial DNA (mtDNA). We summarize advances that have been made in the past 12 years especially in the area of (*a*) accommodating rate variation among sites, among data partitions, and among lineages; (*b*) understanding the information content of data; and (*c*) taking advantage of the relative phylogenetic usefulness of mitochondrial and nuclear genes. We do not include a section on the phylogenetic usefulness of different mtDNA genes because this has been updated for animals by others (e.g., Lin & Danforth 2004, Meyer & Zardoya 2003). Similarly, useful reviews have appeared recently that focus on animal mitochondrial genome evolution (Boore et al. 2005), cytonuclear coevolution (Burger et al. 2003, Rand et al. 2004), mechanisms of gene order rearrangement (Boore 2000), the use of mtDNA in phylogeographic/species-level studies (Funk & Omland 2003), and the population biology of mtDNA (Ballard & Rand 2005). Finally, we include as a web-resource an updated compilation of conserved mtDNA polymerase chain reaction (PCR) primers (see the Supplemental Appendix; follow the Supplemental Material link from the Annual Reviews home page at **http://www.annualreviews.org/**) using the standardized naming system of Simon et al. (1994). This compilation contains 70 new primers that are useful for sequencing large sections of the mitochondrial genome.

### The Beginnings of Molecular Systematics and the Rapid Pace of Change: The Influence of Molecular Technology

In 2003, the world celebrated the 50th anniversary of the discovery of the structure of DNA. Since 1953, DNA sequences have been incorporated into every aspect of biology. The development of molecular technology and subsequent production of data have dictated the direction of molecular phylogenetics. Despite the advances introduced by chain termination sequencing (Sanger et al. 1977), DNA sequencing was difficult and slow before the advent of PCR (Saiki et al. 1985); molecular phylogenetic analysis was therefore largely based on amino acid sequences, immunological distances, DNA-DNA hybridization, allozymes, and mitochondrial DNA restriction site mapping (reviewed in Simon 1991). Before the development of large batteries of conserved PCR primers for mitochondrial DNA (e.g., Kocher et al. 1989, Simon et al. 1994), direct sequencing of RNA was easier than sequencing of DNA, and large data sets of 18S ribosomal RNA (rRNA) accumulated and grew for comparative purposes. Thus was set into motion the collection of a large amount of sequence data

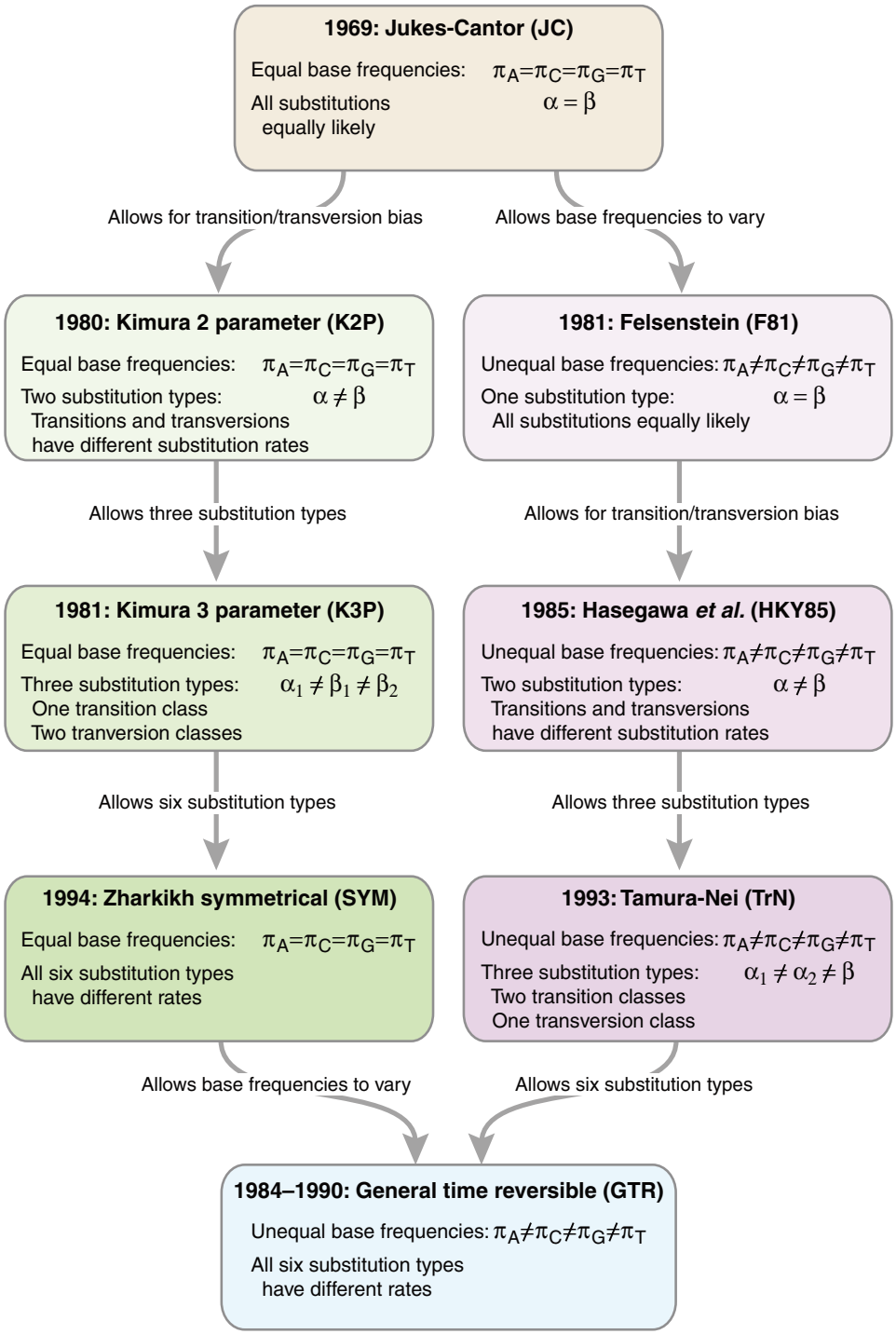
for a severely problematic—thus very interesting—macromolecule simply because of sequencing technology. In the early to mid-1990s, PCR primers and the development of fast and efficient automated sequencing machines greatly increased the rate of collection of DNA data. The sequencing of complete organelle and nuclear genomes has greatly facilitated the development of additional PCR primers and the selection of genes. Despite the promise of nuclear genes (Zhang & Hewitt 2003), mtDNA still remains the most used genome in animal phylogenetics for studies of mid- to late Cenozoic-age divergences because of its faster rate of evolution, ease of sequencing, paucity of visible recombination, and conserved gene content (Caterino et al. 2000, Lin & Danforth 2004, Simon et al. 1994). A greater understanding of mitochondrial evolution (Ballard & Rand 2005, Funk & Omland 2003) allows potential pitfalls in data interpretation to be recognized and avoided.

## The Importance of Phylogenetic Computer Applications

The presence of user-friendly tree-building programs has heavily influenced the choice of phylogenetic methods made by most systematists. The value of model-based methods such as maximum likelihood (ML) became apparent as more was learned about the mechanisms of evolution of DNA sequences, and empiricists and theoreticians began to consider the necessity of accommodating the peculiarities of molecular evolution in increasingly realistic models (Swofford et al. 1996; see **Figure 1**). In the 1980s and 1990s tree-building programs that implemented maximum likelihood facilitated model-based analyses (e.g., Felsenstein 1981; Swofford 1998, beta version in 1993; Yang 1997, beta version 1993). In 2001, the first version of the program, MrBayes, became available (Huelsenbeck & Ronquist 2001). Because these programs allow a strong focus on models of evolution, building realistic models and choosing among them are two of the most active areas of research in systematics today (Posada & Buckley 2004, Sullivan & Joyce 2005). User-friendly programs like Modeltest (Posada & Crandall 1998) have facilitated the selection of models.

In the precursor to this review (Simon et al. 1994), we pointed out that likelihood and spectral-analysis methods "show great promise for phylogenetic analysis but are computationally intensive and currently work well only for a limited number of taxa." For this reason, models of evolution were discussed in terms of distance corrections and parsimony weighting. Between 1996 and 2001, as computers and algorithms picked up speed, likelihood became the method of choice. Bayesian phylogenetic analysis (Huelsenbeck et al. 2001, Larget & Simon 1999, Yang & Rannala 1997) was rapidly embraced once MrBayes became available (Huelsenbeck & Ronquist 2001). The advantage of Bayesian analysis lies in its ability to reveal phylogenetic uncertainty in trees directly constructed using probabilistic models (Holder & Lewis 2003, Huelsenbeck & Imennov 2002). Leache & Reeder (2002) were the first to compare parsimony, likelihood, and Bayesian phylogenetic analyses for a large mitochondrial data set and discuss the comparative advantages of these procedures.

Today, nucleotide sequence data are accumulating faster than they can be analyzed. Better and better models of evolution are being developed. Still, it is not apparent whether current Bayesian tree-building and fast maximum likelihood (e.g., Zwickl



2006) programs that incorporate complex models can handle the very large numbers of taxa to be analyzed in future data sets. One major question is whether large numbers of genes and especially enormous numbers of finely sampled taxa (sequences) can rescue distance analyses that do not make full use of the character information in the data and nonmodel-based methods such as evenly weighted parsimony that ignore complex substitution patterns. The remainder of this review explores substitution patterns and their significant effects on phylogenetic analyses and data interpretation.

**Molecular clock:** the assumption that the rate of molecular substitutions is constant per unit time and can be used to date divergences

## HOW MOLECULES EVOLVE

### Evolution and Weighting of Molecular Data

Our previous review (Simon et al. 1994) describes how model-based corrections and analogous phylogenetic weighting schemes were devised to correct for the fact that nucleotide substitutions at single sites are obscured by later substitutions that can mislead phylogenetic and molecular clock analyses. Beginning with the Jukes-Cantor (1969) model, we traced the parallel development of weighting schemes and models of evolution that relax each of Jukes-Cantor's unrealistic assumptions: (*a*) all bases are found in equal proportions within a sequence, (*b*) every base changes to every other base with equal probability, and (*c*) the rate of substitution at every site is the same.

To incorporate molecular realism, parsimony tree-building methods rely on (*a*) weighting (e.g., Cunningham 1997), or (*b*) conversion to distances, correction with a model of evolution and conversion back to character data using a Hadamard transformation (Penny et al. 1996). Because many proponents of parsimony insist on even weighting of bases (which makes the same unrealistic assumptions as the Jukes-Cantor model) and because complex weighting takes away one of parsimony's greatest advantages relative to maximum likelihood and Bayesian analyses—speed—little research progress has been made in data weighting. So, the discussion below focuses on models of evolution. Although some models will always fit data better than others, data

**Figure 1**

The figure shows models of evolution from simplest (most unrealistic) at top to the most complex (most general) at bottom (modified from Swofford et al. 1996, their figure 11, and Page & Holmes 1998, their figure 5.14). $\alpha =$ transition rate; $\beta =$ transversion rate. All models are symmetrical (probability of changing from base X to base Y is the same as changing from base Y to base X). HKY85 is similar to the Felsenstein 1984 (F84) model (formally described by Kishino & Hasegawa 1989) in that both allow for unequal base frequencies and unequal transition and transversion rates. The general time reversible (GTR) model was developed several times between 1984 and 1990 (Felsenstein 2004), but not implemented until 1993 (e.g., Swofford 1993) for computational reasons. Note that none of the models described above include an accommodation for among-site rate variation (ASRV) but this can be added by attaching a $\Gamma$, invariant sites correction, and/or by partitioning data. Standard ASRV corrections all assume that the pattern of ASRV does not change over time; violation of this assumption is addressed by covarion-like models. Another factor not accommodated by the models shown is correlation among sites. Note also that although some of these models accommodate nucleotide bias, this accommodation assumes that the bias is the same in all taxa.

might not fit any model very well (Bollback 2002). Still, Sullivan & Swofford (2001) convincingly argue that it is better to use a poorly fitting model than no model at all. Posada & Buckley (2004) and Sullivan & Joyce (2005) discuss the need for models in science and phylogenetics, review the properties of molecular evolutionary models, and/or compare and contrast methods of model selection. Below we discuss models from the point of view of the data.

**Among-site rate variation (ASRV):** a ubiquitous property of molecules where different numbers of substitutions per unit time occur at different DNA or amino acid sites

**Among-lineage rate variation (ALRV):** a property of molecules where the number of substitutions per unit time occurring at any given position varies among taxa

**Covarion-like evolution:** nucleotide or amino acid substitutions whose pattern of ASRV varies across lineages; encompasses heterotachy, mosaic evolution, and covarion/covariatide evolution

## Nucleotide and Amino Acid Substitution Models

The models of evolution described in **Figure 1** show the progressive incorporation of realism from top to bottom including accommodation of biased base composition within a sequence and biased substitution patterns from one nucleotide to another. As stated in the legend, none of these models incorporate among-site rate variation (ASRV) per se, but ASRV can be added as an additional parameter(s). Accommodating ASRV has more of an effect on phylogenetic analyses than accommodating nucleotide and base substitutional biases within sequences (Sullivan & Swofford 2001, Yang et al. 1994). Adding ASRV to amino acid evolution models has also resulted in substantial improvement in model fit (e.g., Susko et al. 2003, contra Yang et al. 1994). Among-lineage rate variation (ALRV; covarion-like evolution) is a more complex process to model and, as a result, models of evolution that address these processes are less well developed. Below we discuss among-site and ALRV and their effects on phylogenetic analysis.

## Among-Site Rate Variation

**History.** Ideally, slowly evolving genes would be most useful for deep-level phylogenetics whereas rapidly evolving genes would be necessary for reconstructing recent divergences. Unfortunately, most genes are not well characterized by a single average rate of evolution. With the unraveling of the genetic code between 1961 and 1966, it became immediately obvious that substitutions at different codon and amino acid positions would be accepted at different rates owing to the degeneracy of first and third positions. Shortly thereafter, evolutionary biologists began to examine how this ASRV might affect genetic distances among taxa and phylogenetic analyses based on them (e.g., Fitch & Margoliash 1967). Simon et al. (1994) reviewed early studies that attempted to incorporate ASRV into models of evolution and weighting schemes. They also demonstrated how knowledge of molecular structure and function can help to understand the constraints that create rate variation among sites. Yang (1996) produced an exceptionally complete review of the discovery of ASRV and its incorporation as both discrete rate classes and continuous distributions into usable models of DNA sequence evolution including methods for calculating the $\alpha$-shape parameter of the $\Gamma$-distribution of rates across sites. It is now well established that even genes that are considered to be strongly conserved contain rapidly evolving sites (e.g., Simon et al. 1996), with the converse also being true. Therefore it is inadequate to characterize a gene as fast or slow and expect that categorization to hold across all sites.

**Detrimental effects of ignoring ASRV.** In his 1996 review, Yang summarized the detrimental effects of ignoring ASRV: severe underestimation of genetic distances, incorrect estimation of transition-transversion rate ratios, and confounding of phylogenetic tree-building algorithms. On the population level, Yang (1996) pointed out that the infinite alleles model ignores ASRV, which in turn invalidates Tajima's D statistic for testing neutrality and causes the distribution of pairwise sequence differences to mimic patterns of population expansion. Revell et al. (2005) found that incorrect inferences of rapid cladogenesis early in a group's history due to bias in tree shape was caused by model underparameterization, especially the omission or misestimation of ASRV. Buckley et al. (2001a) and Buckley & Cunningham (2002) showed that ignoring ASRV has a strong effect on estimates of nonparametric bootstrap support. Similarly, Lemmon & Moriarty (2004) found that when ASRV is ignored in simulations model misspecification has a strong effect on Bayesian posterior probability estimates of nodal support. Although Kumar et al. (1993) warned that correcting sequences that differ by less than 5% may result in overparameterization (trading an increased error variance for realism), Yang (1996) argued that this point of view is a misconception. Sullivan & Joyce (2005) present an extensive discussion of the impact of model misspecification and overparameterization.

**SSR:** site-specific rates

**Topology:** the branching order of a phylogenetic tree

**Among-site rate variation in amino acid sequences.** Like nucleotides, amino acids show considerable variability in rate of substitution among sites. Although the earliest studies of ASRV focused on amino acids (e.g., Fitch & Markowitz 1970), later phylogenetic studies of amino acids ignored ASRV. Recently, there has been a rebirth of interest in ASRV and the beneficial effects of its incorporation into models of amino acid evolution (e.g., Susko et al. 2003, 2004). As noted above, ASRV is caused by structural and functional constraints. Mitochondria-specific models of evolution that use information on the probability of particular amino acid replacements (e.g., Adachi & Hasegawa 1996) and secondary structure (e.g., Lió & Goldman 2002) also improve phylogeny construction, especially for deeper-level phylogenetic studies where adaptive shifts in molecules among lineages (see below) are more likely to have taken place.

**Different methods of accommodating among-site rate variation vary in their effects.** ASRV can be addressed by partitioning the data into different rate classes and assigning each rate class its own rate (site-specific rates or SSR models). However, SSR models (where each site in a rate class is assumed to evolve at the same rate) give much lower branch-length estimates than Γ models, invariant sites models, and partitioned-Γ models (**Figure 2**). Although ignoring ASRV may not in all circumstances affect topology, it is more likely to have an effect on nodal support (Buckley et al. 2001a). Buckley & Cunningham (2002) evaluated the effect of different ASRV models using six real data sets for taxa whose relationships are supported strongly by other data. By examining the match of trees constructed from a variety of models to the well-supported trees, they found that SSR models and evenly weighted parsimony performed poorly in recovering the topology and produced lower branch supports than models that incorporated a gamma (Γ) or invariant sites (I) correction.

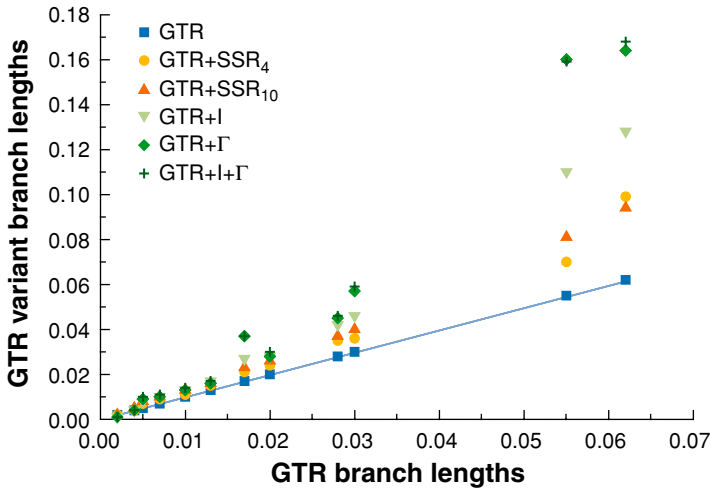



**Figure 2**

Maximum-likelihood estimates of branch lengths under a range of variants of the general time reversible (GTR) substitution model (*colored symbols*). The diagonal line connecting the GTR points on the x- and y-axes illustrates the deviation of the ASRV-corrected branch lengths from the equal-rates estimates. Models of evolution are described in **Figure 1**. Subscripts indicate number of partitions; e.g., in the GTR + $SSR_4$ model the data is partitioned into four SSR classes (first, second, and third positions of proteins plus transfer RNA sites). Redrawn with permission from Buckley et al. 2001a.

## Among-Partition Rate Variation

**APRV:** among-partition rate variation

**Mixture models:** models in which data points are viewed as generated by one of a number of distributions, each contributing to the likelihood

Although partitioning data does not work well via SSR models because of the usually unrealistic assumption that all sites within the partition evolve at the same rate, partitioning strategies that estimate the distribution of ASRV separately for each partition (SSR+Γ) avoid this problem. Yang's PAML program has allowed partitioned models for many years but worked slowly for large numbers of taxa. Recently MrBayes has introduced partitioned models for Bayesian Markov chain Monte Carlo (MCMC) analyses that work well for many taxa. Using partitioned models on combined data with heterogeneous rates addresses many of the concerns about data combinability (e.g., Barker & Lutzoni 2002, Buckley et al. 2002, Bull et al. 1993)—assuming that the same topology (history) underlies different data partitions. Independently modeling data partitions improves branch support and tree likelihood (Brandley et al. 2005, Castoe et al. 2004, Nylander et al. 2004), but must be done with care because partitioning data without properly accommodating among-partition rate variation (APRV) and/or using branch-length priors that are too diffuse can seriously distort branch lengths (Marshall et al. 2006).

## Mixture Models

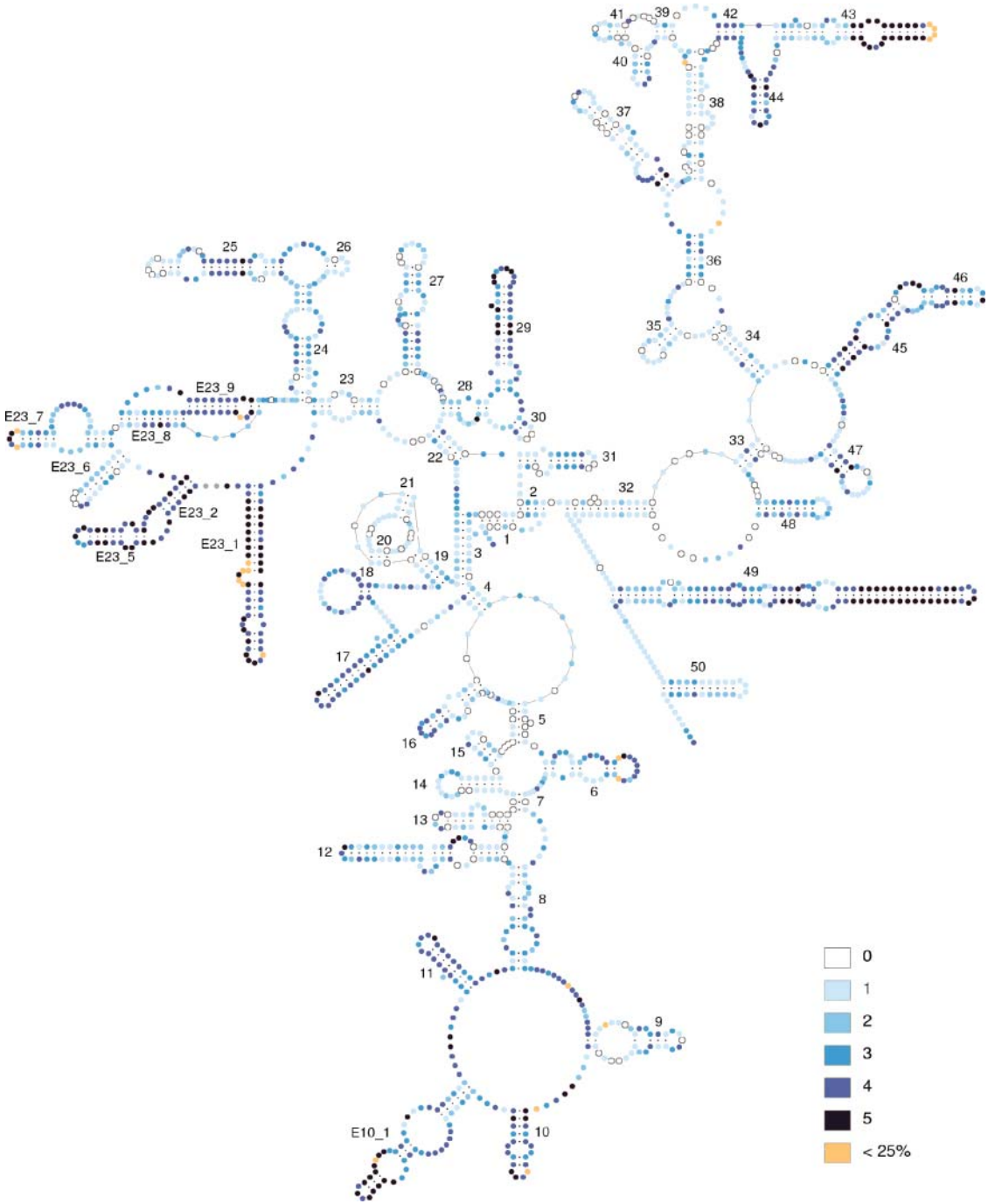
There are many ways any one data set can be partitioned, and the method by which this is done is usually arbitrary. Indeed, for many genes it is not always clear how

best to partition data. For protein coding genes, first, second, and third positions are common partitions; however, within each of these categories there is considerable variation in functional constraint. At third positions, sites can be twofold, threefold, or fourfold degenerate with the fourfold degenerate sites displaying the most freedom to vary (although these sites may be somewhat constrained by codon usage bias). In addition, messenger RNA secondary structure, protein secondary/tertiary structure, and other functional considerations all influence the rate and pattern of evolution. In rRNA, partitioning stems (paired) versus loops (unpaired) regions is tempting, but there is heterogeneity within each of these partitions. Although some authors have partitioned rRNA data into stems versus loops (e.g., Springer & Douzery 1996), this does not make sense in terms of partition variability because some stems are slowly evolving whereas others evolve rapidly (**Figure 3**). Furthermore, within a stem the pattern of variation may differ depending on the identity of the base; tertiary structure and protein interactions also add complexity (Gutell et al. 2002, Hickson et al. 1996, Pagel & Meade 2004). A more informative method of accommodating rate and pattern variation among data subsets is to use mixture models that do not require a priori specification of partitions. In mixture models, characters are viewed as having been generated by one of a number of distributions, and each of these distributions contributes to the likelihood during an analysis. The parameters of each model distribution and the weights assigned to them are estimated from the data. In the end of the analysis, each character can be assigned a probability of membership in each model. Pagel & Meade (2004) developed a Bayesian mixture model and characterized pattern variation in published protein (EF1$\alpha$ and DDC) and 12S small subunit (SSU) mitochondrial rRNA data. For the rRNA data, their program converged on four rate matrices to describe the patterns of variation in the data but these matrices only partially corresponded to stems and loops. In fact, the sites weighted heaviest for one of the matrices were evenly divided between stems and loops. Similarly, for the protein coding genes, although each Q matrix specialized on a particular codon position, each matrix also provided the best fit to some other codon positions. Lartillot & Philippe (2004) developed a mixture model for amino acid sequence evolution. Their use of a Bayesian Dirichlet process prior allowed the association of each amino acid site to a given Q matrix to be determined during the analysis. Using Bayes factors they demonstrated that the mixture model outperformed standard amino acid rate matrices.

**rRNA secondary structure:** the pattern of helices and unpaired regions formed when a single-stranded ribosomal RNA primary sequence folds and pairs with itself

**Covariation:** two bases or two amino acids whose rates of evolution are correlated usually due to functional or structural constraints

**Mixture models do not solve all problems for ribosomal RNA.** In rRNA there is clearly covariation among sites that is related to base-pairing in helices, long-distance base-pairing interactions across domains, and ribosomal-protein-rRNA interactions (e.g., Hickson et al. 1996). In fact, covariation analysis has been particularly important in devising elegant models of rRNA secondary structure that have now been completely verified using experimental methods and X-ray crystallography (Gutell et al. 2002). Accommodating correlation among sites in models of rRNA evolution is important for phylogenetic analysis (e.g., Huelsenbeck & Nielsen 1999, Smith et al. 2004) and needs further study. Finally, rRNA molecules often include large numbers of variable length indels that can cause alignment difficulties and provide information that is difficult to objectively incorporate into phylogenetic reconstruction. In general,



information in indels has been quantified separately and added into phylogenetic analyses later (Kjer et al. 2001, Lutzoni et al. 2000). A growing number of models explicitly describe covariation between bases in RNA molecules (e.g., Smith et al. 2004), but these are only rarely employed in phylogenetic reconstructions (e.g., Kjer 2004).

## Among-Lineage Rate Variation

**Covarions, covariatides, heterotachy.** Fitch & Markowitz (1970) proposed that adaptive shifts in protein function over time would result in a change in the probability of substitution of a particular amino acid site over time (across lineages). Based on earlier ideas by Margoliash & Smith (1965) and Fitch & Margoliash (1967), they proposed the covarion hypothesis, which speculated that a certain proportion of sites in a protein were not free to vary but could become variable if other sites assumed their function. This was later extended to nucleotides and called covariotide evolution (Fitch 1986). Miyamoto & Fitch (1995) reviewed the development of the covarion hypothesis and showed that the distribution of variable and invariant positions is different in seven mammal and seven plant sorbitol dehydrogenase amino acid sequences and thus follows a covarion model. Miyamoto & Fitch (1995) and others (e.g., Steel et al. 2000) pointed out that a constant-sized covarion class is unrealistic because the number of variable positions can differ among lineages. Newer covarion models (Galtier 2001, Tuffley & Steel 1998) do not require this restriction.

Because nucleotide functional/rate shifts occur in rRNA as well as in protein coding genes (Simon et al. 1994, 1996), a more general name for this phenomenon would be helpful. Philippe & Lopez (2001) coined the useful term heterotachy to describe positions that evolve at different rates in different lineages. Earlier this type of evolution had been called "independent and episodic" (Johannes & Berger 1993), "mosaic evolution" (Simon et al. 1996), or "covarion-like evolution" (Lockhart et al. 1998, Lopez et al. 1999). Demonstrations that evolutionary rate of a given position is not always constant throughout time, apart from discussions of codon evolution, include those by Johannes & Berger (1993), Philippe et al. (1996), Simon et al. (1996), Lockhart et al. (1996, 1998), Lopez et al. (1999), and Gaucher et al. (2001).

Functional shifts in molecules cause shifts in the pattern of ASRV and are expected over the course of long-term evolution. Lopez et al. (2002) studied more than 2000 vertebrate cytochrome-b sequences from 32 large monophyletic groups and found

**Figure 3**

Rates of variability of individual nucleotide positions contingent on nucleotide variabilities were based on 500 sequences of species belonging to the eukaryotic crown taxa small subunit rRNA molecules, superimposed on the secondary structure of *Saccharomyces cerevisiae*. The most variable positions are in black, the most conserved in light blue, and invariable positions are in white. Sites containing a nucleotide in *S. cerevisiae* but vacant in more than 75% of sequences, which were not considered for the variability calculations, are indicated in orange. Areas that could not be aligned with confidence are also indicated in orange. All rRNA molecules have similar patterns of variability. Redrawn with permission from Yves Van de Peer (http://www.psb.ugent.be/rRNA/varmaps/Scer_ssu.html).

that all variable positions are heterotachous but that, surprisingly, there was no obvious relationship between variability, function and three-dimensional structure. Gribaldo et al. (2003) found that not all sites in vertebrate hemoglobin show a relationship to structure and function; there is a class of sites that are constant (within groups) but different (among groups) and these CBD sites show "signatures of functional specialization." Lockhart et al. (2000) showed that evolving distributions of variable sites alone provide support for deep-branching patterns in eubacterial phylogenies. If these shifts exhibit convergent evolution, Lockhart and colleagues pointed out, the tree or parts of it could be artifactual (bad covarion evolution). Reassuringly, different genes recovered similar patterns of evolution (good covarion evolution). Misof et al. (2002) showed how the Γ-distributed rate at particular sites in the insect mitochondrial 16S rRNA gene varied extensively among different insect orders, indicating that covariotide evolution operates in mtDNA at deep levels of divergence.

**Nonstationarity:** any process characterized by statistical properties that vary over time; in phylogenetics most often discussed with reference to nucleotide bias

Progress has been made in developing covarion-like models of sequence evolution. In addition to the Tuffley & Steel (1998) and Galtier (2001) models cited above, Yang et al. (2000) created a model that allows selection constraints to change across the sequence. One drawback of this model is that the branches with different selection pressures must be specified a priori. Guindon et al. (2004) created a model that not only allows selection to change across the sequence, but also allows selective constraints to change over time without a priori specification. Recently, Huelsenbeck et al. (2006) applied a Dirichlet process prior in a fully Bayesian approach to model variation in nonsynonymous sites and allow selection to vary across a sequence. These models are close in spirit to the original covarion models in which an amino acid substitution at one position in a gene changes the selective constraints elsewhere.

**Nucleotide bias among lineages.** Shifts in patterns of ASRV can cause changes in the nucleotide bias of taxa across the tree. This is because the nucleotide bias of an organism's genome is most evident at the most variable sites (Simon et al. 1994). Earlier it had been shown that substitutional bias can seriously affect phylogenetic trees (e.g., Lockhart et al. 1992, Weisburg et al. 1989). If patterns of nucleotide bias have changed over lineages, models that assume a stationary distribution of nucleotide bias among taxa can cause systematic error in phylogeny construction. LogDet-type models were designed to incorporate nucleotide bias nonstationarity (Steel 1994). Because ASRV and ALRV (covarion-like evolution) can occur simultaneously the LogDet model should be combined with an invariant sites model because the LogDet does not correct for ASRV. Haddrath & Baker (2001) showed that ratite bird phylogenies based on complete mitochondrial genomes were consistent with traditional expectations only when they were corrected for variation in nucleotide bias among lineages. Similarly, using complete mitochondrial genomes corrected for nonstationarity, Paton et al. (2002) refute the controversial conclusions based on fossils/morphology that modern birds are descended from shorebirds or passerines. Jermiin et al. (2004) review the literature on nucleotide nonstationarity, discuss methods to detect it and, using simulations, examine the effects of nucleotide nonstationarity on tree building.

Although the biasing effects of nucleotide nonstationarity are well known, they are often forgotten. For example, in a recent paper, Rokas et al. (2003) used data

from eight complete yeast nuclear genomes to discover the minimum number of genes necessary to build a robust and well-supported phylogeny. They concluded that approximately 20 genes (average length 1198 nucleotides) were needed and that there were no discernable characteristics of these genes that "predicted the performance" of one over the other. There are several problems with these conclusions. First—trivial but often overlooked—it is not the number of genes that is important but the number of informative sites, and this differs among genes and at different depths in the tree. Second and more important, there are many characteristics of genes that can make them less useful for phylogenetic analysis. One is high levels of ASRV; another is high levels of ALRV. Collins et al. (2005) demonstrated that many of the genes that Rokas et al. (2003) employed contained nonstationary nucleotide frequencies. When these genes were excluded, it was concluded that on average, eight yeast genes were required to recover the underlying phylogeny. Whereas, as Collins and colleagues point out, genes do not come in two classes (stationary and nonstationary), the greater the variation in levels of base compositional bias among lineages, the more problematic the gene is likely to be for phylogenetic reconstruction. A recent paper by Hedtke et al. (2006) demonstrated that the Rokas et al. (2003) phylogeny also suffered from poor taxon sampling leading to long branch problems.

**Codon models.** Codon models, reviewed by Yang (2003), operate at the level of the codon as a unit. They are an advancement over amino acid substitution models that only consider the probability of changing from one particular amino acid to another. Codon models take into account the fact that a switch from one type of codon family (e.g., twofold versus fourfold versus sixfold degeneracy) changes the probability of substitution of individual bases within the codon. Although earlier codon models assumed that the rate of synonymous substitution is constant among sites within genes (e.g., Muse & Gaut 1994) newer codon models reflect the fact that there is significant variability of synonymous rates in the majority of genes. This is observed, for example, in complete mitochondrial genome sequences of 111 animal taxa sampled from 10 disparate clades (F.V. Mannino & S.V. Muse, in review). Because of their computational complexity, codon models are only rarely used for inferring tree topology (Ren & Yang 2005).

## INTERPRETING TREES AND DATA SUPPORT

A thorough knowledge of the properties of molecular data and how they evolve can also aid in the design and intrepretation of phylogenetic studies. Taxon sampling, measuring nodal support, testing alternative phylogenetic hypotheses and attaching time to phylogenies are four areas where understanding molecular evolution can improve the process.

### Taxon Sampling

Theoretical and simulation studies have shown that trees can be very hard to reconstruct when branch lengths are unequal and rates of change vary over the tree,

although this is partly dependent on the method of phylogenetic inference employed (e.g., Felsenstein 1978, Swofford et al. 2001). Adding taxa to a tree in order to break up long branches improves the accuracy of topology estimation by better reconstructing the history of character-state changes, an implicit step in parsimony and likelihood inference. The more nodes in a phylogenetic tree, the more information on character-state change. The result is that more densely sampled subsections of a tree tend to have longer total branch lengths (Venditti et al. 2006). Another factor that can reduce sampling bias is increased sequence length, which also leads to more accurate estimates of branch lengths and tree topology as long as the added nucleotides have properties similar to the original sample. The above observations have led to a debate on whether it is preferable to sample more characters or more taxa in order to increase phylogenetic accuracy (e.g., Pollock et al. 2002, Rosenberg & Kumar 2001, Zwickl & Hillis 2002). The extensive simulation studies of Zwickl & Hillis (2002) and Pollock et al. (2002) showed how adding taxa to a phylogenetic analysis was more effective than adding more characters once a certain sequence length was reached (Hillis et al. 2003). Maximum-likelihood simulations showed consistent improvement with the addition of taxa, probably owing to the improved estimation of model parameters (Pollock & Bruno 2000). However, taxon addition should be done strategically to avoid adding taxa that increase the depth of the tree or that create long branches, which can introduce further biases into the reconstruction (e.g., Mitchell et al. 2000, Poe 2003, Hedtke et al. 2006).

The second debate surrounding the density of taxon sampling concerns the efficacy of evenly weighted parsimony in reconstructing large trees with many taxa. Hillis (1996) suggested that densely sampled phylogenies were actually easier to reconstruct accurately than were sparsely sampled phylogenies, even for simple models of evolution such as evenly weighted parsimony. He showed that highly variable rates of evolution and ASRV were much less of a problem for densely sampled trees than for trees of only a few taxa. Dense sampling of taxa decreases the number of superimposed changes of characters that must be reconstructed along lineages and therefore decreases the reliance on accurate and complex models of evolutionary change. DeBry (2005) summarized the ensuing debate and conducted new simulations, which agreed that overall accuracy for parsimony increases with increased taxon sampling. Similarly, Salamin et al. (2005) showed that evenly weighted parsimony performed quite well in reconstructing large phylogenies. However, although it is inevitable that increased taxon sampling will help to reconstruct superimposed changes, the conclusions about the efficacy of evenly weighted parsimony may be less applicable to empirical sequence data, which tend to fit models less well than simulated data and have more ASRV and more nucleotide bias (e.g., Holder 2001). Empirical data may also show variation in patterns of substitution among lineages as described above.

## Measures of Nodal Support

Measures of nodal support are generally more satisfying than whole tree measures of information content because for most phylogenetic trees some clades are better supported than others. Measures of nodal support provide a useful summary

of how well data support the relationships defined by a tree. Poorly supported relationships are of little use in evolutionary studies other than to illustrate where more data are needed before conclusions can be drawn. Of course, high support values do not mean that a node is accurate, only that it is well supported by the data; model misspecification and taxon sampling can mislead the analysis (e.g., Hedtke et al. 2006).

Currently, the nonparametric bootstrap (Felsenstein 1985) is still the most widely used method for assessing nodal support, despite a long-running debate as to the validity and interpretation of the bootstrap in phylogenetics (e.g., Efron et al. 1996, Felsenstein & Kishino 1993, Hillis & Bull 1993, Holmes 2005, Sanderson 1995). Perhaps the best interpretation is that the bootstrap quantifies the sensitivity of a node to perturbations in the data (Holmes 2005). However, as commonly implemented, the bootstrap gives a biased estimate of accuracy (Hillis & Bull 1993, Holmes 2005), where accuracy is defined as the probability of obtaining a correct phylogenetic reconstruction (Penny et al. 1992). The reason for this bias is related to the complex geometry of tree space and site-pattern space (Efron et al. 1996, Holmes 2005), which is described as follows. All possible site patterns (i.e., sets of sites that show identical states across all taxa) for a given data set can be divided into regions, each one separated by a boundary. Each region has an optimal topology associated with it. An observed data set lies within the sampling distribution of what we can consider to be the truth. Because bootstrap replicates are generated from the observed data rather than the truth, the proportion of replicate data sets that lie within each region can become distorted, which in turn can bias the bootstrap (Sanderson & Wojciechowski 2000). More sophisticated bootstrap techniques are available to correct for this bias (e.g., Efron et al. 1996, Shimodaira 2002); unfortunately, these are rarely implemented to measure nodal support.

The well-known bias of the bootstrap has led researchers to seek other methods of estimating nodal support, and perhaps the most popular alternative is Bayesian posterior probability (e.g., Larget & Simon 1999, Yang & Rannala 1997). The increasing reliance on posterior probabilities as measures of nodal support, as opposed to the bootstrap, has initiated a debate as to the merits of the two approaches (e.g., Huelsenbeck & Rannala 2004, Suzuki et al. 2002). This debate arose from early observations of Bayesian inference in phylogenetics that demonstrated a tendency for posterior probabilities to be more extreme than ML nonparametric bootstrap proportions, although the two tended to be correlated. This observation was made from both empirical (e.g., Buckley et al. 2002, Wilcox et al. 2002) and simulated data (e.g., Cummings et al. 2003, Suzuki et al. 2002, Wilcox et al. 2002). Here we address the following questions: Why are bootstrap proportions and posterior probabilities different? Is this really a problem? If so what can be done about it?

Comparing posterior probabilities and bootstrap proportions is difficult because they represent fundamentally different quantities. A nodal posterior probability is the probability that a given node is found in the true tree, conditional on the observed data, and the model (including both the prior model and the likelihood model). Some researchers argue that posterior probabilities are superior to bootstrap

proportions because the former give a more direct measure of confidence in a node (e.g., Huelsenbeck & Rannala 2004). A bootstrap proportion is based on the concept of resampling, and its exact interpretation depends on how it was calculated, as discussed above. Furthermore, posterior probabilities are calculated by assuming a prior distribution on all model parameters, including the branch lengths and topology, and these priors will influence the posterior in many cases (e.g., Yang & Rannala 2005). This dependence on prior distributions also complicates the comparison of the two measures of support. Yang & Rannala (2005) demonstrated how some prior distributions on branch length can cause nodal posterior probabilities to become extreme. Lewis et al. (2005) also showed how posterior probabilities can be biased if a prior that excludes zero-length branches is applied to data generated from a topology that, in fact, includes polytomies, an observation first made by Suzuki et al. (2002). Lewis et al. (2005) demonstrated that if a polytomy exists but is not accommodated in the prior, resolution of the polytomy will be arbitrary and the nodal support indicated by the posterior probability will appear unusually high compared to ML bootstraps. As with the problems noted by Yang & Rannala (2005), this can be circumvented by applying more appropriate prior distributions or by using reversible jump MCMC to permit internal branches not supported by the data to collapse to polytomies (Lewis et al. 2005).

Another observation from simulated (Huelsenbeck & Rannala 2004, Lemmon & Moriarity 2004, Suzuki et al. 2002) and empirical data (Buckley 2002, Waddell et al. 2002) is how model misspecification affects posterior probabilities relative to bootstrap proportions. The simulations by Huelsenbeck & Rannala (2004) and the empirical study of Buckley (2002) show how posterior probabilities respond in a more extreme fashion to model misspecification than the bootstrap or bootstrap-based topology tests. This problem is likely to be exacerbated by branch-length heterogeneity (Felsenstein and inverse-Felsenstein zone problems) and a high rate of change across the tree, both of which typify many mtDNA data sets. This problem can obviously be rectified by implementing a more complex substitution model; however, there is no guarantee that the models as implemented in available software packages will be sufficient. Because we have little knowledge of the goodness of fit between data and model in typical phylogenetic studies (although goodness of fit tests do exist), we have little idea of the seriousness of the problem of model misspecification in current implementations of Bayesian phylogenetic inference. Finally, failure of convergence of the MCMC algorithm is an underappreciated problem, especially for large data sets. Failure to diagnose a lack of convergence of the algorithm will lead to incorrect posterior probabilities (Huelsenbeck et al. 2002).

Given these issues, which method is best for quantifying phylogenetic support? This is a difficult question to answer because it partly depends on one's philosophical approach to statistical inference. However, if the desired measure of support is the probability that a node is correct given the data set and the model, then the only way to calculate this is by Bayes' theorem. Some researchers have attempted to reconcile Bayesian and bootstrap approaches by merging multiple Bayesian analyses from bootstrapped data sets, the so-called Bayesian bootstrap (Douady et al. 2003,

Waddell et al. 2002). However, the exact statistical interpretation of these values is not at all obvious, and this combination of distinct statistical paradigms has yet to be justified. In terms of practical advice, a review of the current Bayesian phylogenetic literature indicates that much more emphasis needs to be placed on developing more realistic models, checking the effects of the priors, and monitoring the convergence of posterior distributions.

## Tests of Topology

In many phylogenetic studies it is important to assess the information content of the data with respect to the entire tree relative to an alternative hypothesis in addition to understanding support for individual nodes. A variety of tests of topology have been designed to achieve this goal. The currently used tests of topology can be divided into two types: frequentist and Bayesian. The frequentist tests can in turn be divided into parametric tests and nonparametric tests. The most widely used parametric test is the Swofford-Olsen-Waddell-Hillis (SOWH) test (Swofford et al. 1996), which uses parametric bootstrapping to simulate replicate data sets that are in turn used to obtain the null distribution. The SOWH test has been applied in a wide variety of studies ranging from comparative phylogeography (e.g., Carstens et al. 2005b) to deeper phylogenetics (e.g., Rokas et al. 2002). The nonparametric tests use the nonparametric bootstrap to generate replicates that are then used to construct the null distribution. The Kashino-Hasegawa (KH) test (Kishino & Hasegawa 1989) is a nonparametric test designed to compare pairs of topologies selected before a phylogenetic analysis and may become too liberal when the maximum-likelihood topology (selected a posteriori) is tested against another topology (Goldman et al. 2000). The Shimodaira-Hasegawa (SH) test (Shimodaira & Hasegawa 1999) and the approximately-unbiased (AU) test (Shimodaira 2002) simultaneously compare sets of topologies and incorporate more complex bootstrap procedures to correct for the bias associated with multiple comparisons and inclusion of the maximum-likelihood topology. The study by Buckley et al. (2001b) demonstrates the effect of these assumptions on the SH test relative to the KH test. For these reasons, and because of the nature of the null hypotheses employed by the nonparametric tests, the SH and KH tests are generally more conservative than the parametric tests (e.g., Aris-Brosou 2003, Buckley 2002, Goldman et al. 2000). The more explicit reliance on models of evolution by the parametric tests makes them very powerful tests, yet they are also more susceptible to model misspecification (e.g., Buckley 2002, Huelsenbeck et al. 1996, Shimodaira 2002).

Bayesian tests of topology (e.g., Aris-Brosou 2003) are much less commonly implemented than the frequentist tests. The Bayesian tests generally rely on Bayes factors (Kass & Raftery 1995) to compare marginal likelihoods generated under two hypotheses corresponding to different topologies. The use of Bayes factors in testing topologies will likely receive much greater attention in the future (Huelsenbeck & Imennov 2002, Suchard et al. 2005). One example of a Bayesian test of topology is that of Carstens et al. (2005a), who assessed whether posterior distributions of trees contained topologies consistent with a priori demographic hypotheses.

## Attaching Time to Phylogenies

The past several years have seen a renewed interest in methods for attaching time estimates to phylogenetic trees (Arbogast et al. 2002, Welch & Bromham 2005). Early attempts to estimate divergence times when the molecular clock was violated often involved removing taxa with aberrant rates of evolution (Hedges & Kumar 2003), decomposing the procedure into quartets of taxa that conformed to the molecular clock (e.g., Rambaut & Bromham 1998), or fitting local clocks to different regions of the phylogeny (Yoder & Yang 2000). More recently, the focus has shifted to explicitly accounting for rate changes over the tree with a growing emphasis on using models to quantify the uncertainty. The first attempts to correct for changing rates with time were the nonparametric and semiparametric rate-smoothing methods described by Sanderson (1997, 2002). The methods based on explicit models, known as relaxed-clock models, used Bayesian estimation to obtain posterior distributions of node times (Aris-Brosou & Yang 2002, Thorne et al. 1998). For example, the method of Kishino et al. (2001) assumes that the rate of evolution along a descendant branch is a random variable drawn from a log-normal distribution, whose mean is the rate of evolution of the parent branch. This approach has been modified by other researchers who use different distributions in relaxed-clock models (e.g., Aris-Brosou & Yang 2002). Given the expanding range of relaxed-clock models it is becoming increasingly important to justify the use of one model over another. However, model selection procedures are rarely applied to relaxed-clock models. Aris-Brosou & Yang (2002) first applied Bayesian model selection to different relaxed-clock models, and we expect these methods to be more commonly applied in the future, especially as different relaxed-clock models are incorporated into user-friendly software packages. We currently have little information as to how well DNA data sets fit the various relaxed-clock models, although simulation and empirical studies show that these methods can be misleading when the relaxed-clock model deviates strongly from the actual process of changing rates over the tree (Welch et al. 2005), as is likely to be true for comparisons of closely related populations experiencing slightly deleterious mutations (Ho et al. 2005). Another serious source of error in dating studies is the manner in which dates are calibrated. It is desirable to have as many calibration points as possible and new methods that improve the ability to incorporate uncertainty into fossil dates are an important step forward (Yang & Rannala 2006).

A further complication for dating divergences using mitochondrial DNA is model misspecification compounded by the typically rapid rate of evolution even for divergences that are only a few million years old. Buckley et al. (2001b) observed large differences in branch-length estimates among substitution models for a group of New Zealand cicada genera that began to radiate 10 Mya (Arensburger et al. 2004). These results indicate that, even for divergences a few million years old, the substitution model can be very important for obtaining reliable divergence times. Furthermore, if data are partitioned, then APRV must be properly accommodated and suitable branch-length priors employed (Marshall et al. 2006).

Finally, when attaching estimates of divergence time to recent speciation events, the well-known discordance between species and gene divergence times must be

taken into account. Coalescent theory predicts that mtDNA haplotype divergence dates will often predate the speciation event by a substantial amount if ancestral effective population size was large (Edwards & Beerli 2000). Jennings & Edwards (2005) showed how up to 10 loci were required before stable estimates of species divergence times were obtained for three closely related species of Australian finches. Although large numbers of loci are rarely available for most empirical studies, this uncertainty can be adequately captured if coalescent times are accounted for in divergence-time estimation (Edwards & Beerli 2000). Various methods have been described to accommodate this potential bias (e.g., Edwards & Beerli 2000); however, these methods and the relaxed-clock models have yet to be implemented in a single framework of divergence-time estimation.



## THE PHYLOGENETIC USEFULNESS OF MITOCHONDRIAL DNA

### Mitochondrial Genes and Phylogeny

At the species level, mitochondrial gene data are by far the most widely used marker for assessing phylogenetic relationships. They offer some advantages over nuclear data for practical reasons (ease of actually obtaining the sequences, direct comparisons across different studies, and higher levels of variability). In addition, mtDNA genes have faster coalescent times owing to the smaller effective population size of their haploid, maternally inherited genomes. Thus, through genetic drift, their gene trees achieve species-level reciprocal monophyly sooner after speciation than do gene trees generated from nuclear substitutions (Sunnucks 2000). The extensive use of mitochondrial genes in phylogenetic reconstructions (e.g., Caterino et al. 2000) has generated an overwhelmingly greater amount of data for mitochondrial genes compared to nuclear genes for all taxa of Metazoa (e.g., over 800 complete, or nearly complete, mitochondrial genomes were available in GenBank as of July 2006). Historically, the mitochondrial genes most often used for phylogenetic purposes are *co1*, *co2*, *ssu* (small subunit) and *lsu* (large subunit) *rRNA*, *cytb*, and the control region (Caterino et al. 2000, Meyer & Zardoya 2003), but most regions of the mitochondrial genome are similarly useful; Simon et al. (1994) discuss the relative usefulness of the various mitochondrial genes at different levels of divergence (see especially their table 1).

### Comparison of Substitution Rates in Mitochondrial versus Nuclear Genes

It has long been known that mitochondrial genes evolve faster than the majority of genes encoded in the nuclear genome (Brown et al. 1982). Although the number of nuclear genes is much greater, and the variance of the average nonsynonymous substitution rates among them is reasonably expected to be much larger, synonymous substitutions of mitochondrial genes have been empirically estimated to accumulate 1.7–3.4 times as fast as in the most rapidly evolving nuclear genes, and 4.5–9 times as fast if one averages across all nuclear genes studied (Moriyama & Powell 1997). These estimates may be biased, however, because genes chosen for analysis are usually

the most conserved in order to facilitate primer design. Nuclear introns, although faster than nuclear coding regions, are still slower than mtDNA (Zhang & Hewitt 2003). Heterogeneity in substitution rate divergence between mitochondrial and nuclear genes has been observed as a function of the genes and taxa studied (Lin & Danforth 2004). Faster rates of evolution in mitochondrial genes have been related to higher rates of transition mutations (Brown et al. 1982) and stronger constraints in nuclear genes due to selection for codon usage (Moriyama & Powell 1997). Other factors believed to influence the rates of evolution of mitochondrial genes include thermal adaptation, mitochondrial-nuclear interactions, and infection with *Wolbachia* (Ballard & Rand 2005). The faster evolution of mitochondrial genes implies higher levels of multiple substitutions than nuclear genes, especially at synonymous sites (Goto & Kimura 2001, Overton & Rhoads 2004), with an obvious effect on levels of homoplasy when genes are used for phylogenetic inference (Lin & Danforth 2004). Another major difference between nuclear and mitochondrial genes that could influence phylogenetic inference is the greater nucleotide compositional bias of mtDNA, especially in insects where A+T bias can be very extreme (Simon et al. 1994). As a further complication, each of the two strands of mtDNA may exhibit different patterns of base compositional bias, and these patterns can change because of gene rearrangements (Hassanin et al. 2005).

## Performance of Mitochondrial versus Nuclear Genes for Phylogeny

The combination of these factors affects phylogenetic inference and has led to many attempts to evaluate the differential performance of nuclear and mitochondrial genes in phylogenetic reconstructions. Mitochondrial and nuclear genes have often been found to differ significantly in phylogenetic signal (Overton & Rhoads 2004), a pattern that is to be expected at shallow phylogenetic levels, where different genes may present truly different allelic histories (gene trees) owing to as yet incomplete sorting of ancestral mtDNA haplotype polymorphisms under drift. Nuclear genes are suspected to outperform mitochondrial genes in phylogenetic inference when the depth of the tree is such that the nuclear genes possess sufficient variability (Lin & Danforth 2004), whereas more rapidly evolving mitochondrial genes will have experienced more multiple substitutions and associated homoplasy. Obviously, the faster-evolving mitochondrial genes provide more resolving power for the phylogeny of closely related taxa and for phylogeographic and population genetic studies (Avise 2000, Zhang & Hewitt 2003), but are more problematic for resolving the deepest nodes of a phylogenetic tree (distantly related taxa), because of the extreme compositional biases, the asymmetry of transformation-rate matrices, the higher amount of homoplasy, and the higher levels of ASRV (Lin & Danforth 2004, Springer et al. 2001). This has led to the conclusion that mitochondrial genes should be used largely for the phylogenetics of closely related taxa (Cenozoic divergences) and that they require highly parameterized models that correct for some of the best known evolutionary anomalies if they are to be used for the phylogenetic analysis of more ancient divergences (Lin & Danforth 2004). Nevertheless, as discussed earlier, rates of evolution are not the only important parameters; other features, such as the heterogeneity of rates of

variation across sites and across lineages, may be more important. In addition, the use of structural information from the sequence of mtDNA-encoded proteins has proven valuable for the resolution of deep phylogenetic relationships, such as those among different eukaryotic lineages (reviewed in Bullerwell & Gray 2004).

However, mitochondrial genes have a number of advantages over nuclear genes that have stimulated their extensive use for phylogenetic inference at all taxonomic levels. In addition to faster coalescence times (discussed above), these features include haploidy, the absence (or reduced rates) of recombination (Birky 2001), maternal inheritance, and more neutral patterns of evolution. Although some of these features have been challenged by evidence of recombination (Guo et al. 2006, Shao et al. 2005), doubly uniparental inheritance (Passamonti et al. 2003), heteroplasmy with paternal leakage (Bromham et al. 2003), and selective sweeps (Ballard & Rand 2005), these phenomena are likely to have little effect on mitochondrial phylogenetics above the species level, especially if rapid phylogenetic radiations need not be resolved (an expensive proposition anyway). Even assuming that recombination occurs, its frequency is certainly much lower in mtDNA than in the nuclear genome, and it most likely occurs between more similar haplotypes. In terms of phylogenetic utility this means that nuclear genes are more likely to be a mosaic of parts having a different evolutionary history and to be telling a different phylogenetic tale, either because of recombination or because of sorting of ancestral polymorphisms (e.g., Pääbo 2003). The mitochondrial genome is much less affected by this problem, and it tends to evolve as a single, rarely recombinant locus. Nuclear genes are most effective if multiple unlinked loci can be sequenced and if these loci are variable enough to answer the questions posed.

The combination of nuclear and mitochondrial data is useful because incongruence between nuclear and mitochondrial phylogenies can reveal important aspects of species histories such as introgression, hybridization, direct or indirect selection, and incomplete lineage sorting (reviewed by Funk & Omland 2003 and Ballard & Rand 2005). Because mtDNA can introgress faster than nuclear DNA and can sometimes entirely replace the mtDNA of the species invaded (Ballard & Rand 2005, Machado & Hey 2003), the combined use of nuclear traits is an important check. It is difficult to list all the different studies where markers from the two genomes have been used in conjunction, but examples can be found of either perfect congruence (e.g., Overton & Rhoads 2004) or sharp conflict (e.g., Goto & Kimura 2001) of the phylogenetic signal. They have helped to reveal recent hybridization events (e.g., Alexandrino et al. 2005, Buckley et al. 2006, Jordan et al. 2003), and cases of peculiar reproductive systems, such as hybridogenesis and androgenesis (e.g., Mantovani et al. 2001).

## Nuclear Pseudogenes

One serious problem with the use of mitochondrial genes for phylogenetic analysis is the occurrence of copies of mitochondrial genes translocated to the nucleus (numts) where they become pseudogenes (Bensasson et al. 2001, Thalmann et al. 2004, Zhang & Hewitt 1996). Identifying pseudogenes by reading-frame disruption is not sufficient, as functional mitochondrial genes containing translational frameshifts have been observed (see the Supplemental Appendix for discussion;

follow the Supplemental Material link from the Annual Reviews home page at **http://www.annualreviews.org/**). A way to overcoming this problem would be to use gene sequences retrieved from longer sections of the mitochondrion or to take advantage of the fact that the mtDNA is a circle to devise clever long PCR strategies for testing the integrity of the molecule (Thalmann et al. 2004).

## Usefulness of Complete Mitochondrial Genome Sequences

The collection of complete mitochondrial genomes and the analysis of information from all or most of the 13 protein-coding genes have increased dramatically in the past decade (Boore et al. 2005) and the use of this information for phylogenetic purposes has therefore intensified (reviewed in Carapelli et al. 2006 for hexapods, and Meyer & Zardoya 2003 for vertebrates). This has allowed the investigation of the phylogenetic usefulness of complete mitochondrial genomes for deep levels and the testing of different approaches to phylogeny reconstruction (Cook et al. 2005, Delsuc et al. 2003, Hassanin et al. 2005, Nardi et al. 2003). A detailed knowledge of the misleading effects of nucleotide compositional and substitutional biases of mitochondrial protein-coding genes, as well as accelerated rates of evolution owing to genomic rearrangements (e.g., Shao et al. 2003), parasitic lifestyles (e.g., Dowton & Austin 1995), and metabolic rate (e.g., Gillooly et al. 2005), allows the development of more realistic models of DNA evolution and the estimation of well-supported phylogenetic reconstructions (Hassanin 2006).

The gene rearrangements of entire mitochondrial genomes provide another potential source of phylogenetic information, especially for the deepest nodes, because shared gene-order changes are assumed to be evolutionarily rare events (and therefore less likely to occur convergently), and may represent synapomorphic characters useful for reconstructing phylogenetic relationships (Boore et al. 2005, Larget et al. 2005). The best known example was provided by Boore et al. (1998), who suggested an insect-crustacean (Pancrustacea) relationship based on the shared translocation of one transfer RNA (tRNA)-encoding gene, but other examples have been obtained from other taxa (reviewed in Boore et al. 2005). However, the discovery of taxa with extensive variation in gene order (Dowton et al. 2003 for hymenopterans; Shao et al. 2003 for hemipteroids; Scouras et al. 2004 for echinoderms) and convergent gene arrangements (Macey et al. 2004 in amphiesbaenian reptiles) suggests caution when interpreting gene order changes in a phylogenetic framework. Several mechanisms have been proposed by which gene order rearrangements may occur, including intramitochondrial recombination (Dowton & Campbell 2001), and duplication followed by random or nonrandom loss of duplicated genes (Boore 2000, Lavrov et al. 2002).

## CONCLUDING REMARKS

Despite recent challenges, mitochondrial data remain at the forefront of phylogenetic studies at all levels of divergence. When used with care (Ballard & Rand 2005, Funk & Omland 2003), mtDNA is particularly valuable in studies of taxa that have evolved in the past 15 million years, where most nuclear genes show very little change. mtDNA

data have played a major role in studies of molecular evolution that have helped to refine models of evolution and understand how rates and patterns of substitution vary across sequences and over time. The combination of mitochondrial and nuclear data provides useful insights that cannot be obtained with either type of data alone. To facilitate the sequencing of mtDNA, we provide an online appendix of conserved PCR primers that complement those presented in the compilation of Simon et al. (1994).

## SUMMARY POINTS

1. Virtually all nucleotide and amino acid sequence data display ASRV within sequences; accommodating this bias will have more of an effect on phylogenetic analyses than any of the other commonly used substitution model parameters.
2. Mixture models can accommodate among-partition rate and pattern variability without the need to pre-assign partitions.
3. Changes in patterns of ASRV over time (covarion-like evolution) can result in nucleotide bias among taxa that can introduce systematic error if ignored.
4. Many sequences show covarion-like evolution. This is a serious problem for tree building because most models of evolution assume that it does not exist.
5. Measures of nodal support reflect the relative information content of the data for different groupings on the tree. A phylogenetic tree with no support values is meaningless.
6. In Bayesian phylogenetic analyses, it is important to check the effects of the priors and to monitor the convergence of posterior distributions of all parameters.
7. Complete mitochondrial genome sequences contain significant phylogenetic information when analyzed using concatenated nucleotide and amino acid sequences or gene order rearrangements.
8. The combination of nuclear and mitochondrial data is important because incongruence between nuclear and mitochondrial phylogenies can reveal important aspects of species histories.

## FUTURE RESEARCH DIRECTIONS

1. The further development of models of rRNA evolution that incorporate correlation among sites, variable length insertions and deletions, and strong differences in rate variation among sites.
2. The development of better models of covarion-like evolution and continued development of user-friendly programs that incorporate the newest models of evolution.

3. Further exploration of the effects of priors and the convergence of posterior distributions in Bayesian phylogenetic analyses.
4. The exploration of new methods for model selection.

## ACKNOWLEDGMENTS

We thank the following people for information, helpful discussions, or comments on all or part of the manuscript: Jeff Boore, Tim Collins, Dan Funk, Francesco Nardi, Mark Pagel, Paul Lewis, David Hillis, John Huelsenbeck, Mark Holder, Pete Lockhart, Hervé Philippe, Dave Rand, Ziheng Yang, Joe Felsenstein, Spencer Muse, Andrew Roger, Brad Schaffer, Jack Sullivan, and Yves Van de Peer. For unselfishly devoting countless hours designing important computer programs that allow progress in the kind of molecular systematics studies presented here we thank Keith Crandall, Alexei Drummond, Joe Felsenstein, Mark Holder, John Huelsenbeck, Daniel Huson, Sudhir Kumar, Paul Lewis, David & Wayne Maddison, Rod Page, Mark Pagel, David Posada, Fredrik Ronquist, David Swofford, Jeff Thorne, Ziheng Yang, Derrick Zwickl, and others. We also thank Eliza Jewitt and Virge Kask for help with the illustrations. Funding from the following grants aided in the production of this manuscript: NSF DEB 0089946 (C.S.), NSF DEB -04-22386 (C.S., T.B.), NSF DEB 05-29679 (C.S.) and funds from the University of Connecticut (C.S.) and Victoria University of Wellington Visiting Professorial Fellowship (C.S.); the New Zealand Foundation for Research, Science and Technology (T.B.), the New Zealand Marsden Fund (T.B., C.S.), the University of Siena (F.F.), and the Italian Ministry for University and Research (F.F.); and an NSERC of Canada Discovery Grant (A.T.B.).

## LITERATURE CITED

Adachi J, Hasegawa M. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* 42:459–68

Alexandrino J, Baird SJ, Lawson L, Macey JR, Moritz C, Wake DB. 2005. Strong selection against hybrids at a hybrid zone in the *Ensatina* ring species complex and its evolutionary implications. *Evol. Int. J. Org. Evol.* 59:1334–47

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu. Rev. Ecol. Syst.* 33:707–40

Arensburger P, Buckley TR, Simon C, Moulds M, Holsinger KE. 2004. Biogeography and phylogeny of the New Zealand cicada genera (*Hemiptera: Cicadidae*) based on nuclear and mitochondrial DNA data. *J. Biogeogr.* 31:557–69

Aris-Brosou S. 2003. How Bayes tests of molecular phylogenies compare with frequentist approaches. *Bioinfomatics* 19:618–24

Aris-Brosou S, Yang Z. 2002. The effects of models of rate evolution on estimation of divergence dates with a special reference to the metazoan 18S rRNA phylogeny. *Syst. Biol.* 51:703–14

Avise JC. 2000. *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard Univ. Press
Ballard JWO, Rand DM. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. *Annu. Rev. Ecol. Evol. Syst*. 36:621–42
Barker FK, Lutzoni FM. 2002. The utility of the incongruence length difference test. *Syst. Biol*. 51:625–37
Bensasson D, Zhang X, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol*. 16:314–21
Birky CWJ. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet*. 35:125–48
Bollback JP. 2002. Bayesian model adequacy and choice in phylogenetics. *Mol. Biol. Evol*. 19:1171–80
Boore JL. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In *Comparative Genomics*, ed. D Sankoff, JH Nadeau, pp. 133–47. Netherlands: Kluwer
Boore JL, Lavrov D, Brown WM. 1998. Gene translocation links insects and crustaceans. *Nature* 393:667–68
Boore JL, Macey JR, Medina M. 2005. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol*. 395:311–48
Brandley MC, Schmitz AS, Reeder TW. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst. Biol*. 54:373–90
Bromham L, Eyre-Walker A, Smith NH, Maynard Smith J. 2003. Mitochondrial Steve: paternal inheritance of mitochondria in humans. *Trends Ecol. Evol*. 18:2–4
Brown WM, Prager EM, Wang A, Wilson AC. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol*. 18:225–39
Buckley TR. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Syst. Biol*. 51:509–23
Buckley TR, Arensburger P, Simon C, Chambers GK. 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Syst. Biol*. 51:4–18
Buckley TR, Cordeiro M, Marshall DC, Simon C. 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). *Syst. Biol*. 55:411–25
Buckley TR, Cunningham CW. 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. *Mol. Biol. Evol*. 19:394–402
**Buckley TR, Simon C, Chambers GK. 2001a. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths and bootstrap support. *Syst. Biol*. 50:67–86**
Buckley TR, Simon C, Shimodaira H, Chambers GK. 2001b. Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (*Maoricicada*) using multiple comparison tests of tree topology. *Mol. Biol. Evol*. 18:223–34
Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol*. 42:384–97

**Show that different methods of correcting for ASRV have varying effects on branch lengths and nodal support.**

Bullerwell CE, Gray MW. 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. *Curr. Opin. Microbiol.* 7:528–34

Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: Anything goes. *Trends Genet.* 19:709–16

Carapelli A, Nardi F, Dallai R, Frati F. 2006. A review of molecular data for the phylogeny of basal hexapods. *Pedobiologia* 50:191–204

Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, Sullivan J. 2005a. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution* 59:1639–52

Carstens BC, Demboski JR, Good JM, Brunsfeld SJ, Sullivan J. 2005b. The evolutionary history of the northern Rocky Mountain mesic forest ecosystem. *Evolution* 59:1639–52

Castoe TA, Doan TM, Parkinson CL. 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. *Syst. Biol.* 53:448–59

Caterino MS, Cho S, Sperling FAH. 2000. The current state of insect molecular systematics: a thriving tower of Babel. *Annu. Rev. Entomol.* 45:1–54

**Demonstrates that nucleotide bias nonstationarity has a strong effect on the phylogenetic usefulness of genes.**

**Collins T, Fedrigo O, Naylor GJP. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. *Syst. Biol.* 54:493–500**

Cook CE, Yue Q, Akam M. 2005. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proc. R. Soc. London Ser. B* 272:1295–304

Cummings MP, Handley SA, Myers DS, Reed RL, Rokas A, Winka K. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* 52:477–87

Cunningham CW. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Syst. Biol.* 46:464–78

DeBry RW. 2005. The systematic component of phylogenetic error as a function of taxonomic sampling under parsimony. *Syst. Biol.* 54:432–40

Delsuc F, Phillips MJ, Penny D. 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?" *Science* 301:1482d

Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* 20:248–54

Dowton M, Austin AD. 1995. Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. *J. Mol. Evol.* 41:958–65

Dowton M, Campbell NJH. 2001. Intramitochondrial recombination—is it why some mitochondrial genes sleep around? *Trends Ecol. Evol.* 16:269–71

Dowton M, Castro LR, Campbell SL, Bargon SD, Austin AD. 2003. Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction. *J. Mol. Evol.* 56:517–26

Edwards SV, Beerli P. 2000. Perspective: gene divergence, population divergence, and the variation in coalescence time in phylogeographic studies. *Evolution* 54:1839–54

Efron B, Halloran E, Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 93:13429–34

Felsenstein J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27:401–10

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368–76

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–91

Felsenstein J. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer. 664 pp

Felsenstein J, Kishino H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst. Biol.* 42:193–200

Fitch WM. 1986. The estimate of total nucleotide substitutions from pairwise differences is biased. *Philos. Trans. R. Soc. London Ser. B* 312:317–24

Fitch WM, Margoliash E. 1967. A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome *c* as a model case. *Biochem. Genet.* 1:65–71

Fitch WM, Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem. Genet.* 4:579–93

Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: Frequency, causes, consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* 34:397–423

Galtier N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol. Biol. Evol.* 18:866–73

Gaucher EA, Miyamoto MM, Benner SA. 2001. Function-structure analysis of proteins using covarion-based evolutionary approaches: elongation factors. *Proc. Natl. Acad. Sci. USA* 98:548–52

Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc. Natl. Acad. Sci. USA* 102:140–45

Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49:652–70

Goto SG, Kimura MT. 2001. Phylogenetic utility of mitochondrial *COI* and nuclear *Gpdh* genes in *Drosophila*. *Mol. Phylogenet. Evol.* 18:404–22

Gribaldo S, Casane D, Lopez P, Philippe H. 2003. Functional divergence prediction from evolutionary analysis: a case study of vertebrate hemoglobin. *Mol. Biol. Evol.* 20:1754–59

**Guindon S, Rodrigo AG, Dyer KA, Huelsenbeck JP. 2004. Modeling the site-specific variation of selection patterns along lineages. *Proc. Natl. Acad. Sci. USA* 35:12957–62**

Guo X, Liu SJ, Liu Y. 2006. Evidence for recombination of mitochondrial DNA in triploid crucian carp. *Genetics* 172:1745–49

**Created a covarion-like model that allows selective constraints to change over time without a priori specification of the branches on which the changes occur.**

**Gutell RR, Lee JC Cannone JJ. 2002. The accuracy of ribosomal RNA comparative structure models. *Curr. Opin. Struct. Biol.* 12:301–10**

Provides elegant, proven, rRNA secondary-structure diagrams expandable in the PDF.

Haddrath OP, Baker AJ. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc. R. Soc. London Ser. B* 268:939–45

Hassanin A. 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitutions. *Mol. Phylogenet. Evol.* 38:100–16

**Hassanin A, Léger N, Deutsch J. 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* 54:277–98**

Studied the effects of gene translocations on strand-specific mutational bias and developed a phylogenetic method to accommodate reversed mutational constraints.

Hedges SB, Kumar S. 2003. Genomic clocks and evolutionary timescales. *Trends Genet.* 19:200–6

Hedtke SM, Townsend TM, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst. Biol.* 55:522–29

Hickson RE, Simon C, Cooper A, Spicer GS, Sullivan J, Penny D. 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* 13:150–69

Hillis DM. 1996. Inferring complex phylogenies. *Nature* 383:130–31

Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182–92

Hillis DM, Pollock D, McGuire JA, Zwickl DJ. 2003. Is sparse taxon sampling a problem for phylogenetic inference? *Syst. Biol.* 52:124–26

Ho SY, Phillips MJ, Drummond AJ, Cooper A. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Mol. Biol. Evol.* 22:1355–63

**Holder M, Lewis PO. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4:275–84**

Provides an excellent overview of the advantages, disadvantages, and functioning of different phylogenetic methods.

Holder MT. 2001. *Using a complex model of sequence evolution to evaluate and improve phylogenetics methods*. PhD thesis. Univ. Texas, Austin. 165 pp.

Holmes S. 2005. Statistical approach to tests involving phylogenetic trees. In *Mathematics of Evolution and Phylogeny*, ed. O Gascuel, pp. 91–121. New York: Oxford Univ. Press

Huelsenbeck JP, Hillis DM, Nielsen R. 1996. A likelihood ratio test of monophyly. *Syst. Biol.* 45:546–58

Huelsenbeck JP, Imennov NS. 2002. Geographic origin of human mitochondrial DNA: accommodating phylogenetic uncertainty and model comparison. *Syst. Biol.* 51:155–65

Huelsenbeck JP, Jain S, Frost SWD, Kosakovsky Pond SL. 2006. A Dirichlet process model for detecting positive selection in protein-coding DNA sequences. *Proc. Natl. Acad. Sci. USA* 103:6263–68

Huelsenbeck JP, Larget B, Miller RE, Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51:673–88

Huelsenbeck JP, Nielsen R. 1999. Effect of nonindependent substitution on phylogenetic accuracy. *Syst. Biol.* 48:317–28
Huelsenbeck JP, Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under complex and simple substitution models. *Syst. Biol.* 53:904–13
Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–55
Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–14
Jennings WB, Edwards SV. 2005. Speciational history of Australian grass finches (*Poephila*) inferred from 30 gene trees. *Evol. Int. J. Org. Evol.* 59:2033–47
Jermiin L, Ho SYW, Ababneh F, Robinson J, Larkum AWD. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.* 53:638–43
Johannes GJ, Berger EG. 1993. Domains within the mammalian ornithine decarboxylase messenger RNA have evolved independently and episodically. *J. Mol. Evol.* 36:555–67
Jordan S, Simon C, Polhemus D. 2003. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). *Syst. Biol.* 52:89–109
Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In *Mammalian Protein Metabolism*, ed. N Munro, pp. 21–132. New York: Academic
Kass RE, Raftery AE. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–95
Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* 29:170–79
Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* 18:352–61
Kjer K. 2004. Aligned 18S and insect phylogeny. *Syst. Biol.* 53:506–14
Kjer KM, Blahnik RJ, Holzenthal RW. 2001. Phylogeny of Trichoptera (Caddisflies): characterization of signal within multiple datasets. *Syst. Biol.* 50:781–816
Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, et al. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci.USA* 86:6196–200
Kumar S, Tamura K, Nei M. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.0. University Park, Pennsylvania: Pennsylvania State University
Larget B, Simon DL. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–59
Larget B, Simon DL, Kadane JB, Sweet D. 2005. A Bayesian analysis of metazoan mitochondrial genome rearrangements. *Mol. Biol. Evol.* 22:486–95
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21:1095–1109
Lavrov DV, Boore JL, Brown WM. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangement: duplication and nonrandom loss. *Mol. Biol. Evol.* 19:163–69

Leache A, Reeder T. 2002. Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Syst. Biol.* 51:44–68

Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol.* 53:265–77

**Demonstrates that reversible jump MCMC analysis can improve estimation of posterior probabilities for polytomies.**

**Lewis PO, Holder MT, Holsinger KE. 2005. Polytomies and Bayesian phylogenetic inference. *Syst. Biol.* 54:241–53**

Lin C-P, Danforth BN. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Mol. Phylogenet. Evol.* 30:686–702

Lió P, Goldman N. 2002. Modeling mitochondrial protein evolution using structural information. *J. Mol. Evol.* 54:519–29

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. *J. Mol. Evol.* 34:153–62

Lockhart PJ, Huson D, Maier U, Fraunholz MJ, Van de Peer Y, et al. 2000. How molecules evolve in Eubacteria. *Mol. Biol. Evol.* 17:835–38

Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc. Natl. Acad. Sci. USA* 93:1930–34

Lockhart PJ, Steel MA, Barbrook AC, Huson DH, Charleston MA, Howe CJ. 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol.* 15:1183–88

Lopez P, Casane D, Philippe H. 2002. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19:1–7

Lopez P, Forterre P, Philippe H. 1999. The root of the tree of life in the light of the covarion model. *J. Mol. Evol.* 49:496–508

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.* 49:628–51

Macey JR, Papenfuss TJ, Kuhel JV, Fourcade HM, Boore JL. 2004. Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. *Mol. Phylogenet. Evol.* 33:22–31

Machado CA, Hey J. 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. London Ser. B* 270:1193–202

Mantovani B, Passamonti M, Scali V. 2001. The mitochondrial cytochrome oxidase II gene in *Bacillus* stick insects: ancestry of hybrids, androgenesis, and phylogenetic relationships. *Mol. Phylogenet. Evol.* 19:157–63

Margoliash E, Smith E. 1965. Structural and functional aspects of cytochrome *c* in relation to evolution. In *Evolving Genes and Proteins*, ed V Bryson, HJ Vogel. New York: Academic. 221 pp.

Marshall DC, Simon C, Buckley TR. 2006. Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. *Syst. Biol.* In press

Meyer A, Zardoya R. 2003. Recent advances in the (molecular) phylogeny of vertebrates. *Annu. Rev. Ecol. Evol. Syst.* 34:311–38

Misof B, Anderson CL, Buckley TR, Erpenbeck D, Rickert A, Misof K. 2002. An empirical analysis of mt 16S rRNA covarion-like evolution in insects: site-specific rate variation is clustered and frequently detected. *J. Mol. Evol.* 55:460–69

Mitchell A, Mitter C, Regier JC. 2000. More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Syst. Biol.* 49:202–24

Miyamoto MM, Fitch WM. 1995. Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol.* 12:503–13

Moriyama EN, Powell JR. 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol.* 45:378–91

Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitutions rates, with application to the chloroplast genome. *Mol. Biol. Evol.* 11:1139–51

Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F. 2003. Hexapod origins: monophyletic or polyphyletic? *Science* 299:1887–89

Nylander JAA, Ronquist F, Huelsenbeck JPP, Nieves-Aldrey JL. 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53:47–67

Overton LC, Rhoads DD. 2004. Molecular phylogenetic relationships based on mitochondrial and nuclear gene sequences for the Todies (*Todus*, Todidae) of the Caribbean. *Mol. Phylogenet. Evol.* 32:524–38

Pääbo S. 2003. The mosaic that is our genome. *Nature* 421:409–12

Page RDM, Holmes EC. 1998. *Molecular Evolution. A Phylogenetic Approach*. Oxford, UK: Blackwell Sci. 352 pp.

**Pagel M, Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequences or character-state data. *Syst. Biol.* 53:571–81**

**Describes a general mixture model within a Bayesian MCMC framework that can simultaneously detect rate and pattern heterogeneity.**

Passamonti M, Boore JL, Scali V. 2003. Molecular evolution and recombination in gender-associated mitochondrial DNAs of the Manila clam *Tapes philippinarum*. *Genetics* 164:603–11

Paton T, Haddrath O, Baker AJ. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. *Proc. R. Soc. London Ser. B* 269:839–46

Penny D, Hendy MD, Lockhart PJ, Steel MA. 1996. Corrected parsimony, minimum evolution and Hadamard conjugations. *Syst. Biol.* 45:593–603

Penny D, Hendy MD, Steel MA. 1992. Progress with evolutionary trees. *Trends Ecol. Evol.* 7:73–79

Philippe H, Lecointre G, Van Le HL, Le Guyader H. 1996. A critical study of homoplasy in molecular data with the use of a morphologically based cladogram. *Mol. Biol. Evol.* 13:1174–86

Philippe H, Lopez P. 2001. On the conservation of protein sequences in evolution. *Trends Biochem. Sci.* 26:414–16

Poe S. 2003. Evaluation of the strategy on long-branch subdivision to improve the accuracy of phylogenetic methods. *Syst. Biol.* 52:423–28

Pollock DD, Bruno WJ. 2000. Assessing an unknown evolutionary process: effect of increasing site-specific knowledge through taxon addition. *Mol. Biol. Evol.* 17:1854–58

Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* 51:664–71

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53:793–808

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–18

Rambaut A, Bromham L. 1998. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* 15:442–48

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol. Evol.* 19:645–53

Ren F, Yang Z. 2005. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. *Syst. Biol.* 54:808–18

Revell LJ, Harmon LJ, Glor RE. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Syst. Biol.* 54:973–83

Rokas A, Nylander JAA, Ronquist F, Stone GA. 2002. A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. *Mol. Phylogenet. Evol.* 22:206–19

Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804

Rosenberg MS, Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proc. Natl. Acad. Sci. USA* 98:10751–56

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, et al. 1985. Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–54

Salamin N, Hodkinson TR, Savolainen V. 2005. Towards building the tree of life: a simulation study for all angiosperm genera. *Syst. Biol.* 54:183–96

Sanderson MJ. 1995. Objections to bootstrapping phylogenies: a critique. *Syst. Biol.* 44:299–320

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14:1218–31

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19:101–9

Sanderson MJ, Wojciechowski MF. 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-*Astragalus* (Leguminosae). *Syst. Biol.* 49:671–85

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–67

Scouras A, Beckenbach K, Arndt A, Smith MJ. 2004. Complete mitochondrial genome DNA sequence for two ophiuroids and a holothuroid: the utility of protein gene sequence and gene maps in the analyses of deep deuterostome phylogeny. *Mol. Phylogenet. Evol.* 31:50–65

Shao R, Dowton M, Murrell A, Barker SC. 2003. Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. *Mol. Biol. Evol.* 20:1612–19

Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M. 2005. Novel mitochondrial gene content and gene arrangement indicate illegitimate intermtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. *J. Mol. Evol.* 60:764–73

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51:492–508

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16:1114–16

Simon C. 1991. Molecular systematics at the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from enzymatically amplified DNA. In *Molecular Techniques in Taxonomy*, ed. GM Hewitt, AWB Johnston, JPW Young, H57:33–71. New York: Springer Verlag, NATO Advanced Studies Institute

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87:651–701

Simon C, Nigro L, Sullivan J, Holsinger K, Martin A, et al. 1996. Large differences in substitutional pattern and evolutionary rate of 12S ribosomal RNA genes. *Mol. Biol. Evol.* 13:923–32

Smith AD, Lui TWH, Tillier ERM. 2004. Empirical models for substitution in ribosomal RNA. *Mol. Biol. Evol.* 21:419–27

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, et al. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 18:132–43

Springer MS, Douzery E. 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. *J. Mol. Evol.* 43:357–73

Steel MA. 1994. Recovering a tree from the leaf colorations it generates under a Markov model. *Appl. Math. Lett.* 7:19–24

Steel MA, Huson D, Lockhart PJ. 2000. Invariable sites models and their use in phylogeny reconstruction. *Syst. Biol.* 49:225–32

Suchard MA, Weiss RE, Sinsheimer JS. 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. *Biometrics* 61:665–73

Sullivan J, Joyce P. 2005. Model selection in phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 36:445–66

**Sullivan J, Swofford DL. 2001. Should we use model-based methods for phylogenetic inference when we know assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Syst. Biol.* 50:723–29**

Sunnucks P. 2000. Efficient genetic markers for population biology. *Trends Ecol. Evol.* 15:199–203

Susko E, Field C, Blouin C, Roger AJ. 2003. Estimation of rates-across-sites distributions in phylogenetic substitution models. *Syst. Biol.* 52:594–603

Susko E, Inagaki Y, Roger AJ. 2004. On inconsistency of the neighbor-joining, least squares, minimum evolution estimation when substitution processes are incorrectly modeled. *Mol. Biol. Evol.* 21:1629–42

Suzuki Y, Glazko GV, Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. USA* 99:15138–43

**Show that ASRV has more of an effect than other model parameters on the accuracy of phylogenetic trees.**

Swofford DL. 1993/1998. *PAUP*: Phylogenetic Analysis Using Parsimony* (**and Other Methods*). Sunderland, MA: Sinauer

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In *Molecular Systematics*, ed. DM Hillis, CM Moritz, BK Mable, pp. 407–514. Sunderland, MA: Sinauer

**Clearly explain the long-branch attraction problem and how it affects different phylogenetic tree-building methods.**

**Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO, Rogers JS. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* 50:525–39**

Thalmann O, Hebler J, Poinar N, Pääbo S, Vigilant L. 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Mol. Ecol.* 13:321–35

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15:1647–57

Tuffley C, Steel M. 1998. Modeling the covarion hypothesis of nucleotide substitution. *Math. Biosci.* 147:63–91

Venditti C, Meade A, Pagel M. 2006. Detecting the node-density artifact in phylogeny reconstruction. *Syst. Biol.* 55:637–43

Waddell PJ, Kishino H, Ota R. 2002. Very fast algorithms for evaluating the stability of ML and Bayesian phylogenetic trees from sequence data. *Genome Inform.* 13:82–92

Weisburg WG, Giovannoni SJ, Woese CR. 1989. The *Deinococcus* and *Thermum* phylum and the effect of ribosomal RNA composition on phylogenetic tree construction. *Syst. Appl. Microbiol.* 11:128–34

Welch JJ, Bromham L. 2005. Molecular dating when rates vary. *Trends Ecol. Evol.* 20:320–27

Welch JJ, Fontanillas E, Bromham L. 2005. Molecular dates for the "Cambrian explosion": the influence of prior assumptions. *Syst. Biol.* 54:672–78

Wilcox TP, Zwickl DJ, Heath TA, Hillis DM. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* 25:361–71

Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analysis. *Trends Ecol. Evol.* 11:367–72

Yang Z. 1997. PAML: a program for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13:555–56

Yang Z. 2003. Adaptive molecular evolution. In *Handbook of Statistical Genetics*, ed. DJ Balding, M Bishop, C Cannings, pp. 229–54. New York: Wiley. 2nd ed.

Yang Z, Goldman N, Friday AE. 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Mol. Biol. Evol.* 11:316–24

Yang Z, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–49

Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol. Biol. Evol.* 14:717–24

Yang Z, Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.* 54:455–70

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using fossil calibrations with soft bounds. *Mol. Biol. Evol.* 23:212–26

Yoder AD, Yang Z. 2000. Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* 17:1081–90

Zhang D-X, Hewitt GM. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol. Evol.* 11:247–51

Zhang D-X, Hewitt GM. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol. Ecol.* 12:563–84

Zwickl DJ. 2006. GARLI version 0.94. **http://www.zo.utexas.edu/faculty/antisense/Garli.html**

Zwickl DJ, Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51:588–98

## RELATED RESOURCE

Sanderson MJ, Shaffer HB. 2002. Troubleshooting molecular phylogenetic analyses. *Annu. Rev. Ecol. Syst.* 33:49–72

Annual Review of Ecology, Evolution, and Systematics

Volume 37, 2006

# Contents

**Indexes**

Supplemental Material: Annu. Rev. Ecol. Syst. 2006.
doi: 10.1146/annrev.ecolsys.37.091305.110018
Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA
Simon, C., T.R. Buckley, F. Frati, J.B. Stewart and A.T. Beckenbach

## Supplement 1: List of Conserved Primers for Mitochondrial Genomes

Below is a list of primers that have proven useful for amplifying and sequencing fragments in mitochondrial genomes of insects. This list is expanded and modified from the earlier list provided by Simon et al. (1994). At the time of publication of that review, the number of complete mitochondrial sequences available for arthropods and other Metazoa was quite limited. The past decade has seen a rapid expansion of available sequences, allowing considerable refinement of the primers in the list. None of the primers from the previous review are included here. Many of the primers in this list, however, are modifications of the most useful primers from the earlier review. These primers have been extensively tested with insects, but there has been only limited testing outside of the insects, and modifications may be appropriate for some groups.

Although the primers have been designed primarily for insects, some have proven useful for a much wider range of taxa. Nearly all of the primers align with homologous regions in many Metazoa as shown in the following primer list, and should prove useful (perhaps with modification) for a wide array of animals.

### *Ancestral Arthropod Genome Arrangement*

The first arthropod mitochondrial genome determined was from *Drosophila yakuba* (Clary and Wolstenholme 1985). The sequence coded nine of the 13 protein coding genes and 14 of the 22 tRNA genes on one of the strands. This strand is referred to as the *majority* coding strand (Simon et al. 1994). The other four protein coding genes, eight tRNA genes and both rRNA genes are coded on the opposite strand (*minority* strand). As additional complete arthropod sequences were reported, it became evident that the genome arrangement in *D. yakuba* is widespread among arthropods, and probably represents the ancestral arrangement for the phylum. Therefore, the *D. yakuba* sequence serves as a convenient standard for labeling of primers, with respect to both position and orientation. It should be understood, however, that the actual position of these primers will vary among taxa, depending on the lengths of the genes, the number of non-coding nucleotides present and especially the arrangement of genes in each taxon. Any gene rearrangement will change the position of the primer, and rearrangements involving inversions will alter the orientation as well.

### *Numbering System*

In keeping with the system established by Simon et al. (1994), the primers are labelled according to their positions in the *D. yakuba* complete sequence (Clary and Wolstenholme 1985). The first two or three characters specify the gene: C1 - C3 (cytochrome oxidase subunits), N1 - N6, N4L (NADH dehydrogenase subunits), CB (cytochrome b) and A6, A8 (ATPase subunits); SR and LR specify small and large ribosomal subunits and TX specify the tRNA genes, where X is the one letter designation for the amino acid decoded. TL1 specifies tRNA-Leu (CUN); TL2 specifies tRNA-Leu (UUR); TS1 specifies tRNA-Ser (UCN) and TS2 specifies tRNA-Ser (AGN).

The first character after the dash specifies the strand, either the ma**j**ority coding strand (J) or the mi**n**ority coding strand (N) in the ancestral gene arrangement. We avoid the use of "forward" and

"reverse" since these directions depend on the orientation of the genes. The ribosomal genes, genes for nad1, nad4, nad4l and nad5, and some of the tRNA genes are coded on the minority strand in the ancestral gene arrangement. In those cases, the 'N' primers are "forward" primers. The number following the 'N' or 'J' is the position of the 3' nucleotide of the primer in the *D. yakuba* genome.

Not only does this system unambiguously identify the location and direction of each primer, it provides additional information. To amplify any region in a genome having the ancestral arthropod gene arrangement, or within any gene (regardless of gene arrangement), simply pair a 'J' primer with a higher numbered 'N' primer. The main exception is for amplifying the A+T rich region, where a 'J' primer in SR should be paired with an 'N' primer at the beginning of the sequence. The difference between the numbers, plus the lengths of the primers themselves gives an estimate of the length of the amplified product (again, with the exception of the A+T rich region). These estimates tend to be quite accurate for amplifications within single genes, but may vary when amplifying across gene boundaries. This approach must be modified if the genome has rearrangements in the region of interest (see below).

***Degeneracy***

Most of the primers in this list are degenerate at one or more sites. We use the IUPAC (International Union of Pure and Applied Chemistry) single letter designations: N = [A,C,G and T]; B = not (A) [= C, G and T]; D = not (C); H = not (G); V = not (U or T); R = purine [A and G]; Y = pyrimidine [C and T]; K = [T and G]; M = [A and C]; S = [C and G]; and W = [A and T]. These designations are recognized by companies that make oligonucleotides.

In the list below, all primers are aligned with sequences from a wide range of taxa, concentrating on insects, but including other arthropods, other invertebrates and two vertebrates. The taxa were chosen to give a broad representation, without over-emphasis on any one group. For a few primers it was not possible to identify the homologous region for more distant taxa. In those cases, those taxa are omitted from the alignment.

In the list, dots indicate a match to the residue in the primer at that site. For mismatches, the nucleotide actually present in the sequence is shown. For degenerate primers, alignments are shown twice, side-by-side. The grouping on the left shows the sequence matches taking degeneracy into account. On the right, the actual nucleotides present in each sequence at degenerate sites are shown.

In practice, there appears to be a trade-off between degeneracy and quality of the PCR results. Non-degenerate primers tend to be more specific for the intended gene, and often give better results. But they may fail if there are mismatches in critical positions. Highly degenerate primers, on the other hand, are usually more versatile, but may prime indiscriminately, giving poor results. For phylogenetic surveys, it may be desirable to reduce the degeneracy based on sequence determined for the group of interest. The primers may be refined simply by sequencing the region surrounding the primer for representatives of the desired taxa (by using bracketing primer pairs), and reducing or eliminating degeneracy where possible.

***Genomic Map***

At the beginning of the list, a map of the probable ancestral arthropod genome is shown, linearized between the A+T rich region and tRNA-Ile. This map is not to scale, although in general, larger genes are shown longer than smaller ones. In particular, the tRNA genes comprise a much smaller proportion of the genome than is indicated. The map is divided into seven sections, from about 1.5 to 2.5 kb each.

Each primer in the list is indicated on this map. Those priming the majority coding strand are shown above each section of the map; those on the minority coding strand are shown below. It is hoped that this map will simplify choosing primer pairs appropriate for any region of the genome.

The only major gene where no primers are given is *nad6*. In insects, the only place in the gene where the G-C content is high enough to define a usable primer is not well conserved. While it is often possible to construct taxon-specific primers within this gene, we have been unable to develop versatile primers in *nad6*. It is usually necessary to jump this region by using a 'J' primer within *nad4* or *nad4l*, paired with an 'N' primer in *cytb* (CB-N11010 is most reliable).

When working with taxa where the gene arrangement is not yet known, primer pairs chosen within a gene should work, assuming the sequence matches are adequate. Information about gene organization can be obtained by using primers that bridge gene junctions. Priming across gene junctions can fail for either of two reasons: one or more critical mismatches are present in one or both primers, or the genes are rearranged. If the genes are rearranged, the primers may be too far apart or may even point the wrong directions. One strategy for examining gene junctions in taxa with unknown gene arrangements is to amplify and sequence internally for as many of the genes as possible, then determine the best primers for bridging across them. In the list below, multiple primers within most of the larger genes are provided and can be used for this purpose.

Among arthropods, the most common rearrangements involve tRNA movements. The primers based on tRNA sequences are reasonably well conserved (see alignments, below), but if the tRNA has moved, the primer will not prime from the expected place. In the list below, the primers based on tRNAs are mostly those with relatively stable positions within the arthropod genomes.

***Functional Genes and Pseudogenes***

In most cases, the primers in the list will amplify the correct, functional mitochondrial sequences from extracts of total genomic DNA, but complications can arise from time to time. One complication is the occasional transfer of mitochondrial sequences to the nucleus.

Prior to the invention of the polymerase chain reaction, study of mitochondrial sequences required a number of steps, including shotgun cloning of restriction enzyme digested DNA extract (usually partially purified for mitochondrial sequence), followed by screening with mitochondrial sequence probes. These probes identified candidate clones based on overall sequence similarity to the probe sequence. The studies often pulled clones that were clearly derived from non-functional sequences, and evidently represent recent transfer of copies of the mitochondrial sequence to the nucleus. Non-functional copies are often referred to as nuclear encoded mitochondrial sequences (numt's). Their presence can complicate analyses of mitochondrial sequences.

PCR has many advantages over the earlier techniques. One is that amplification requires an exact (or very close) match to specific sequences conserved in the functional genes, not just an overall sequence similarity of the entire sequence. Mutations occurring in pseudogene copies are more likely to reduce, rather than enhance, the match to the primer binding sites. Nonetheless, a recent transfer (numt) might retain primer binding site sequences and amplify along with the functional mitochondrial gene. It is therefore important to distinguish between functional sequences and numt's.

For protein coding genes, a common approach is to look for open reading frames (ORF's) that correspond to the expected genes. In the vast majority of cases, functional mitochondrial protein coding genes translate as ORF's when using the appropriate genetic code. Recently it has become apparent that some evidently functional mitochondrial genes have had their reading frames disrupted by single nucleotide insertions. These cases include the *nad3* gene of many birds and turtles (Mindell et al. 1998) and the *cytb* gene of some ants (Beckenbach et al. 2005). Their function most likely involves a compensatory frameshift during translation.

What if your mitochondrial protein coding sequence is not an ORF? We suggest the following steps:

1. Recheck the sequence by examining the scans. Automated sequencing often mis-reads scans where there is a long series of a single nucleotide, or if one trace is much stronger than the others. Most anomalies disappear at this step.

2. Sequence the other strand. The two strands are completely different sequences (complementary), except in the case of inverted repeats and palindromic sequences. If a particular sequence causes problems for some sequencing chemistries, this step will identify those problems. This should be done as a matter of routine with automated sequencing, as it will identify many problems with the sequencing.

3. Reamplify and re-sequence the region using a different set of conserved primers. Amplifying a longer piece increases the chance of avoiding nuclear copies. The primer list included here allows amplification of most regions with different conserved primer pairs. Re-sequencing the region will either confirm or correct the anomaly.

If the disrupted ORF is not a PCR or sequencing artefact, it may still be non-functional. Testing for this possibility depends on the nature of the study. For complete mitochondrial genome sequences, the sequence is present in the context of the entire genome. Unless there is a duplication of the gene, it is probably functional. For phylogenetic surveys, if two or more species show the same anomaly, the well-established evolutionary constraints on protein coding genes can provide evidence of their continued functional requirement. Both Mindell et al. (1998) and Beckenbach et al. (2005) used this approach to argue that the frameshifted sequences they observed are functional.

***Amplifying the Control Region (A+T rich region) in Arthropods***

The control region is the least well understood part of the mitochondrial genome. Among arthropods, it is highly variable in length, ranging from only 73 base pairs (bp) in the wallaby louse (Shao, et al. 2001) to 4.6 kilobase pairs (kb) in *Drosophila melanogaster* (Lewis et al. 1994), and 13 kb or more in some beetles (Boyce et al. 1989). Other properties of the control region make it more challenging to amplify, sequence and analyze than the coding portion of the genome. Arthropod control regions often contain tandem repeats, making proper alignment and sequencing difficult. As there is little conservation at the sequence level, versatile primers evidently cannot be designed within the region. And because of the relatively low sequence complexity (consisting primarily of only two nucleotides, A+T), even designing taxon-specific primers may be difficult. Primers consisting of only A+T require low anneal temperatures and tend to bind promiscuously. We offer a few suggestions for studying the control region in arthropods:

1. Assume long PCR techniques will be required, unless there is prior knowledge that the region in the taxa of interest is relatively short. The long PCR techniques have the additional advantage in that they use high quality enzymes that are less likely to fall off during extension. These enzymes may be necessary even if the control region is relatively short (see 4, below).

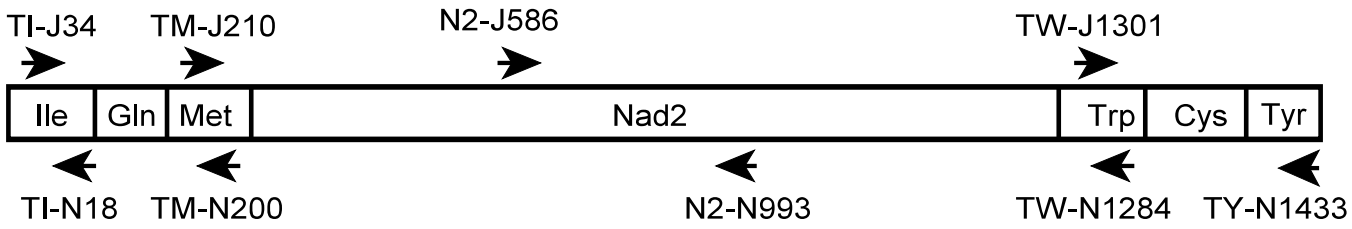
2. Amplify enough of the coding region at each end to insure that the sequence is actually mitochondrial in origin, not just some random AT rich sequence. Since there is no coding information evident within the region, there is no other way to verify its source.

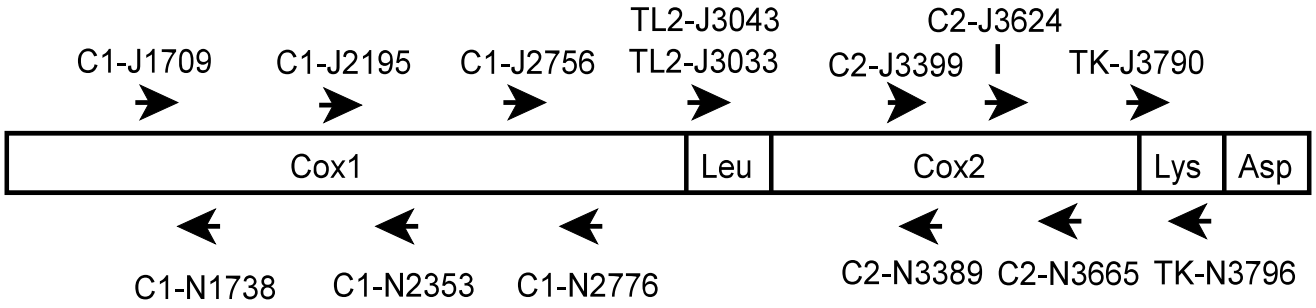
3. Design primers that match exactly to the bracketing coding regions, and have high and similar anneal temperatures.

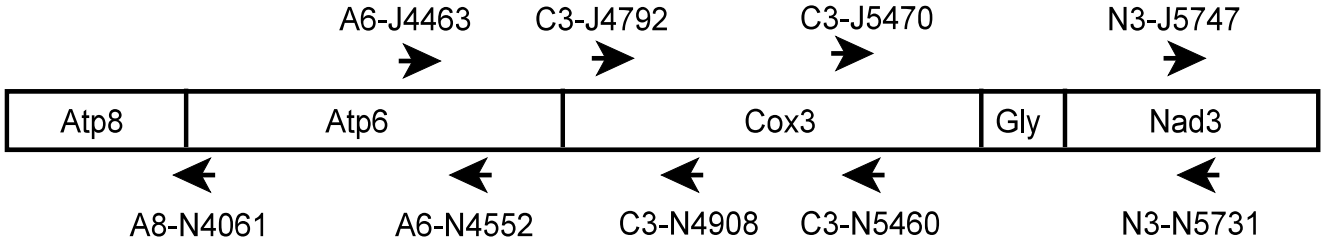
4. Use a relatively low extension temperature, between 65 and 68°C (Stephen Cameron, personal communication). If the polymerase falls off prior to completion, the nascent strand can melt from the template at the usual extension temperature of 72°C. If your PCR primers will anneal between 65 and 68°C, a two step amplification may be appropriate, combining the anneal and extension steps.

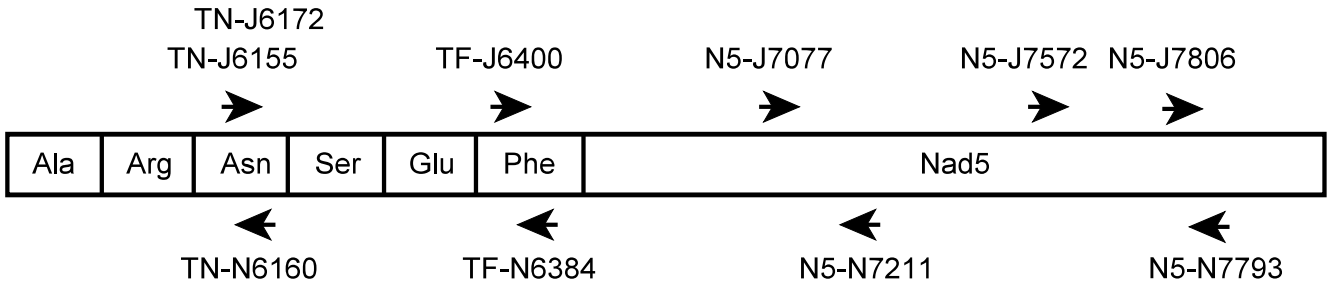
**Table 1**. List of sequences used for primer comparisons.

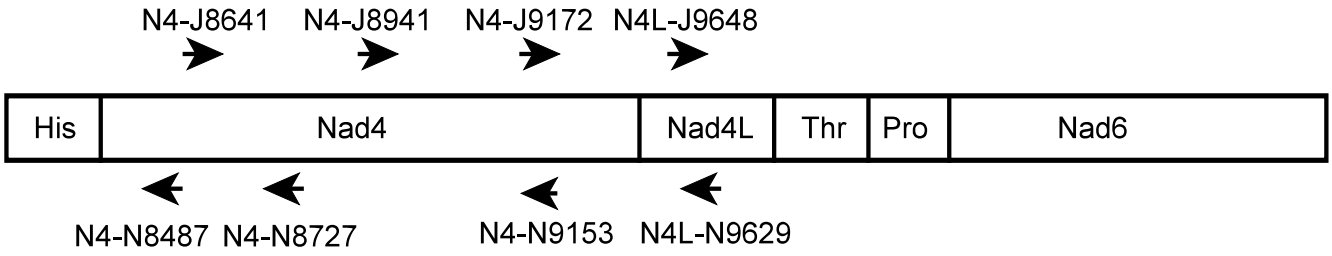
| Organism | GenBank RefSeq or Accession | Reference |
|---|---|---|
| Insecta: | | |
| Diptera - True flies | | |
| *Drosophila yakuba* - "fruitfly" | NC_001322 | Clary and Wolstenholme 1985 |
| *Anopheles gambiae* - mosquito | NC_002084 | Beard et al. 1993 |
| Lepidoptera - Butterflies and Moths | | |
| *Bombyx mori* - silkmoth | NC_002355 | Lee et al. unpublished |
| Coleoptera - Beetles | | |
| *Crioceris duodecimpunctata* - asparagus beetle | NC_003372 | Stewart and Beckenbach 2003 |
| *Tribolium castaneum* - flour beetle | NC_003081 | Friedrich and Muqim 2003 |
| Hymenoptera - Wasps, ants, bees, etc. | | |
| *Apis mellifera* - honeybee | NC_001566 | Crozier and Crozier 1992 |
| *Perga condei* - sawfly | AY787816 | Castro and Dowton 2005 |
| Megaloptera - Dobsonflies | | |
| *Corydalis cornutus* - dobsonfly | - | Beckenbach and Stewart unpublished |
| Hemiptera - True bugs | | |
| *Philaenus spumarius* - meadow spittlebug | NC_005944 | Stewart and Beckenbach 2005 |
| *Triatoma dimidiata* - kissing bug | NC_002609 | Dotson and Beard 2001 |
| Orthoptera - Grasshoppers, locusts | | |
| *Locusta migratoria* - migratory locust | NC_001712 | Flook et al. 1995 |
| Plecoptera - Stoneflies | | |
| *Pteronarcys princeps* - giant stonefly | NC_006133 | Stewart and Beckenbach 2006 |
| Phthiraptera - Lice | | |
| *Heterodoxus macropus*- wallaby louse | NC_002651 | Shao et al. 2001 |
| Other Hexapods: | | |
| Collembola | | |
| *Tetrodontophora bielanensis* | NC_002735 | Nardi et al. 2001 |
| Other Arthropods: | | |
| Crustacea | | |
| *Daphnia pulex* - water flea | NC_000844 | Crease 1999 |
| *Artemia franciscana* - brine shrimp | NC_001620 | Perez et al. 1994 |
| Vertebrates: | | |
| Mammalia | | |
| *Homo sapiens* | AC_000021 | Anderson et al. 1981 |
| Reptilia | | |
| *Chrysemys picta* - painted turtle | NC_002073 | Mindell et al. 1999 |
| Other Invertebrates: | | |
| Mollusca | | |
| *Katharina tunicata* - chiton | NC_001636 | Boore and Brown 1994 |
| *Cepaea nemoralis* - snail | NC_001816 | Terrett et al. 1996 |
| Annelida | | |
| *Lumbricus terrestris* - earthworm | NC_001673 | Boore and Brown 1995 |

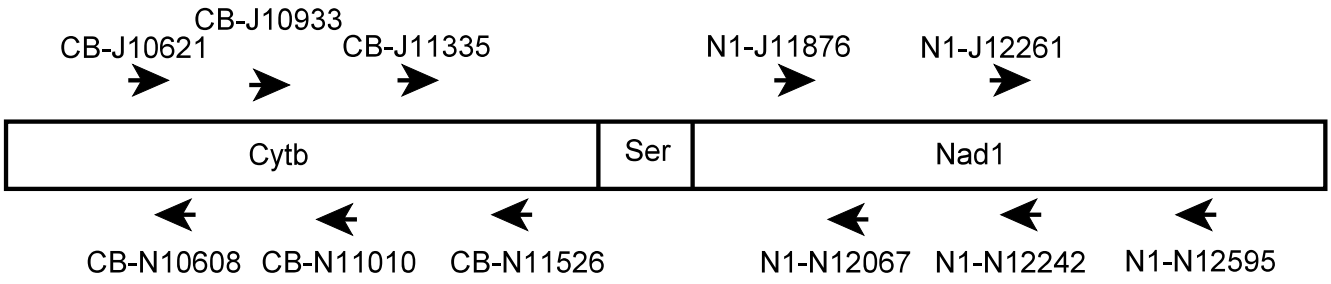

TI-J34
TM-J210
N2-J586
TW-J1301
Ile
Gln
Met
Nad2
Trp
Cys
Tyr
TI-N18
TM-N200
N2-N993
TW-N1284
TY-N1433

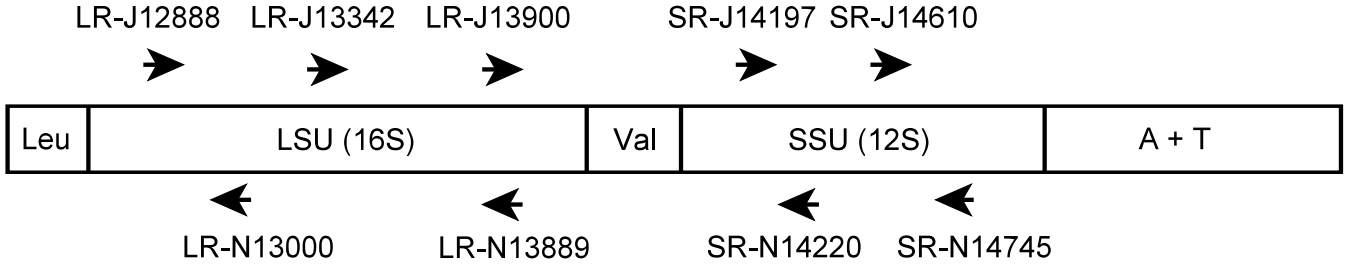

C1-J1709
C1-J2195
C1-J2756
TL2-J3043
TL2-J3033
C2-J3399
C2-J3624
TK-J3790
Cox1
Leu
Cox2
Lys
Asp
C1-N1738
C1-N2353
C1-N2776
C2-N3389
C2-N3665
TK-N3796


A6-J4463
C3-J4792
C3-J5470
N3-J5747
Atp8
Atp6
Cox3
Gly
Nad3
A8-N4061
A6-N4552
C3-N4908
C3-N5460
N3-N5731


TN-J6172
TN-J6155
TF-J6400
N5-J7077
N5-J7572
N5-J7806
Ala
Arg
Asn
Ser
Glu
Phe
Nad5
TN-N6160
TF-N6384
N5-N7211
N5-N7793


N4-J8641
N4-J8941
N4-J9172
N4L-J9648
His
Nad4
Nad4L
Thr
Pro
Nad6
N4-N8487
N4-N8727
N4-N9153
N4L-N9629


CB-J10933
CB-J10621
CB-J11335
N1-J11876
N1-J12261
Cytb
Ser
Nad1
CB-N10608
CB-N11010
CB-N11526
N1-N12067
N1-N12242
N1-N12595


LR-J12888
LR-J13342
LR-J13900
SR-J14197
SR-J14610
Leu
LSU (16S)
Val
SSU (12S)
A + T
LR-N13000
LR-N13889
SR-N14220
SR-N14745

```
TI-N18                10        20         without degeneracy:
(22mer)   5'-TATCCTATCAARRTAAYCCTTT  5'-TATCCTATCAARRTAAYCCTTT
Drosophila   ..C.................     ..C........GG...C.....
Mosquito     ..C.................     ..C........GG...T.....
Silk moth    ...T......G.........     ...T......GAA...T.....
Asp. beetle  ...T................     ...T.......AA...T.....
Flour beetle ...T................     ...T.......AA...T.....
Honeybee     ...TT.........T.T....     ...TT......AA..TTT....
Dobsonfly    ..C............T.....     ..C........GA..TT.....
Froghopper   ...TAC..............     ...TAC.....AA...T.....
Kissing bug  ..............G......     ...........GA..GC.....
Locust       ....................     ...........GA...C.....
Stonefly     ..........C.........     ..........CGA...T.....
Louse        ....T...............     ....T......AA...C.....
Collembola   ..C.......G....C.....     ..C.......GGA..CT.....
Daphnia      ..C............G.....     ..C........GG..GC.....
Artemia      ..CAG......C........     ..CAG......CG...T.....
Human        ..CT.............T....     ..CT.......AG...CT....
Turtle       C.C.................     C.C........GG...C.....
Chiton       ..CAAC.....T...G...G..     ..CAAC.....TG..GC..G..
Snail        AT.GAC.....C.AT.G..G.C     AT.GAC.....CAAT.G..G.C
Earthworm    A.CGTC....G...T....G..     A.CGTC....GAG.T.T..G..
```

TI-N18 is a 22 base primer with two-fold degeneracy at three sites. This primer works well with many insects, and may be used to amplify the control region of insects having the ancestral gene arrangement. TI-N18 overlaps TI-J34 by 17 bases at the 3' end.

```
TI-J34                10        20             without degeneracy:
(26mer)   5'-GCCTGATA-AAAAGGRTTAYYTTGATA  5'-GCCTGATA-AAAAGGRTTAYYTTGATA
Drosophila   ........-..................     ........-......G...CC......
Mosquito     ........-..................     ........-......A...CC......
Silk moth    ......AT--..........C.....     ......AT--.....A...TTC.....
Asp. beetle  .......T--..................     .......T--.....A...TT......
Flour beetle ......---..................     ......---......A...TT......
Honeybee     .T.....--.....A.A.........     .T.....--.....AAA..TT......
Dobsonfly    ......AT-T......A.........     ......AT-T.....AA..TC......
Froghopper   ......A---................G     ......A---.....A...TT.....G
Kissing bug  .......--.......C.........     .......--......GC..TC......
Locust       .......T--..................     .......T--.....G...TC......
Stonefly     .......--............G.....     .......--......A...TCG.....
Louse        .......T-...................     .......T-......G...TT......
Collembola   ........-.......G....C.....     ........-......AG..TCC.....
Daphnia      .......T--......C...........     .......T--.....GC..CC......
Artemia      ........-...........G......     ........-......A...CG......
Human        .T.....--.....A.............     .T.....--.....AG...CT......
Turtle       ......ATTT..................     ......ATTT.....G...CC......
Chiton       ...G.G.G--..C...C...A.....G     ...G.G.G--..C..GC..CA.....G
Snail        ...G.CGG---.C..C.AT.G.....G     ...G.CGG---.C..C.ATTG.....G
Earthworm    ...G...G--..C....A...C....G     ...G...G--..C..A.A.CTC....G
```

TI-J34 is a 26 base primer with two-fold degeneracy at three sites. It works well with most insects. TI-J34 overlaps TI-N18 by 17 bases at the 3' end. The presence of a one or two nucleotide gap, or an extra nucleotide, has surprisingly little effect on the binding ability of this primer.

```
TM-N200                10        20           without degeneracy:
(24mer)   5'-ACCTTTATAARTGGGGTATGARCC  5'-ACCTTTATAARTGGGGTATGARCC
Drosophila   ........................     ..........A..........A..
Mosquito     ........................     ..........G..........A..
Silk moth    T........TT..A..........     T........TT..A.......G..
Asp. beetle  .........CT.............     .........CT..........A..
Flour beetle ..T..C..T..CA...........     ..T..C..T.GCA........A..
Honeybee     TAA.....CG.CA...........     TAA.....CGACA........A..
Dobsonfly    .........T.C............     .........TAC.........A..
Froghopper   .TT.C.......A...........     .TT.C.....A.A........A..
Kissing bug  ..T.C.....T.A...........     ..T.C.....T.A........A..
Locust       .....C..................     .....C....A..........A..
Stonefly     ....C...................     ....C.....A..........A..
Louse        .......--CT.............     .......--CT..........A..
Collembola   ..T.AC...G.A........G...     ..T.AC...GAA........GG..
Daphnia      TAACC..........C.......     TAACC.....A.....C....A..
Artemia      ..TC.AT-C.T.............     ..TC.AT-C.T..........G..
Human        ...AAC..TTTC........G...     ...AAC..TTTC........GG..
Turtle       ...AAC..TTT.........G...     ...AAC..TTT.........GG..
Chiton       .AA.C.CAT...T...........     .AA.C.CAT.G.T........A..
Snail        .GT.ACCAT.T.T..T........     .GT.ACCAT.T.T..T.....G..
Earthworm    TATC.CT.TTTC............     TATC.CT.TTTC.........A..
```

TM-N200 is a 24 base primer with two-fold degeneracy at two sites. It works with many insects and can be easily adapted for specific groups with a few changes in the middle of the primer. It is particularly useful for amplifying the control region. TM-N200 overlaps TM-J210 by 11 bases.

```
TM-J210                10        20
(24mer)   5'-AATTAAGCTACTAGGTTCATACCC
Drosophila   ............G...........
Mosquito     ..C.......T.G...........
Silk moth    .TAA.....TT.G..C.......T
Asp. beetle  ..G.......T.G...........
Flour beetle ...C......T.............
Honeybee     ...A......AC............
Dobsonfly    ...A.....CA.G...........
Froghopper   ...A.......A............
Kissing bug  ...A......T.............
Locust       ...C......A.G...........
Stonefly     ...A......A.G...........
Louse        .T........T.G...........
Collembola   ...A.-......G..CC.......
Daphnia      ...A......G.G.......G...
Artemia      ............G..C........
Human        ...A.-....TCG..CC.......
Turtle       .C.A.-....T.G..CC.......
Chiton       GT.A.-....T.G...........
Snail        GTC.......G.G..C.....A..
Earthworm    .T.C......G.G...........
```

TM-J210 is a non-degenerate 24 base primer that gives excellent results with most insects. Paired with either TW-N1284 or TY-N1433, it can be used to amplify the entire *nad2* gene in many insects. It could be made more versatile by replacing the ‘A’ at position 13 with either a ‘G’ or ‘R’. It should be shortened at the 3' end if used with moths. TM-J210 overlaps TM-N200 by 11 bases.

```
N2-J586                10        20     without degeneracy:
(20mer)   5'-CCATTTCAYTTYTGATTYCC  5'-CCATTTCAYTTYTGATTYCC
Drosophila   ..T.................     ..T.....T..T.....T..
Mosquito     ...................     ........C..C.....T..
Silk moth    ...................     ........T..T.....T..
Asp. beetle  ..........A........     ........C..A.....C..
Flour beetle .....C.............     .....C..T..T.....C..
Honeybee     ..C...........A.AAT     ..C.....T..T...A.AAT
Dobsonfly    ..C.................     ..C.....C..C.....T..
Froghopper   .....C...A.A...GC...     .....C..TA.A...GCT..
Kissing bug  .....C.............     .....C..T..C.....T..
Locust       ..............C....     ........C..T...C.T..
Stonefly     ..C..C.............     ..C..C..T..T.....C..
Louse        ..T.......GA......TT     ..T.....T.GA.....TTT
Collembola   ..T...........C.G..     ..T.....T..T...C.G..
Daphnia      ..C..C...AGA..GG...T     ..C..C..CAGA..GG.C.T
Artemia      ..G..C...A.A.....AGT     ..G..C..TA.A.....AGT
Human        ..C...........G....     ..C.....C..C...G.C..
Turtle       ................A..     ........C..T.....A..
Snail        ..C..A........GG..GT     ..C..A..C..T..GG.TGT
Chiton       ..............GG....     ........T..T..GG.T..
Earthworm    ..C.G...........A...     ..C.G...T..C....AC..
```

N2-J586 is a 20 base primer with two-fold degeneracy at three sites. This region is one of the few places in the gene with sufficient GC content, and conservation at the nucleotide level, to make a reasonably versatile primer. It works for many insects, but needs modification for some groups.

```
N2-N993                10        20          without degeneracy:
(23mer)   5'-GGTAAAAATCCTAAAAATGGNGG  5'-GGTAAAAATCCTAAAAATGGNGG
Drosophila   ......................     ....................A..
Mosquito     ................A.....     ................A..G..
Silk moth    ....T...A..C.T.........     ....T...A..C.T......T..
Asp. beetle  ..C.....C........G.....     ..C.....C........G..A..
Flour beetle ..A.GG.......G.........     ..A.GG.......G......T..
Dobsonfly    ..A.G...................     ..A.G...............G..
Froghopper   ..............T.GA.....     ..............T.GA..A..
Kissing bug  ...........A.T...A.....     ...........A.T...A..G..
Locust       ....G........GT.T......     ....G........GT.T...T..
Stonefly     ..A.G......A.T.........     ..A.G......A.T......A..
Louse        AT......A.....T..A.....     AT......A.....T..A..A..
Collembola   ATGGC...C.....T.GG.....     ATGGC...C.....T.GG..G..
Daphnia      ..A.G...C.....G..G.....     ..A.G...C.....G..G..T..
Artemia      AAGGGCT.A.AC..G..G.....     AAGGGCT.A.AC..G..G..G..
Human        ..C.....G..GGTT.GC.....     ..C.....G..GGTT.GC..G..
Turtle       ..AGT.......GTT.G......     ..AGT.......GTT.G...G..
Snail        .AGGC..C.AAA...T..C....     .AGGC..C.AAA...T..C.C..
Chiton       CCA.....G..A.GT..A.....     CCA.....G..A.GT..A..G..
Earthworm    ..C.TG......GT...A.....     ..C.TG......GT...A..G..
```

N2-N993 is a 23 base primer with four-fold degeneracy at one site. It works with a wide range of insects, and could be easily modified for specific groups. There is no region in the honeybee *nad2* gene that aligns with this region. Sequence for the other hymenopteran, the sawfly, is not available for this region.

```
TW-N1284              10        20          without degeneracy:
(23mer)   5'-ACARCTTTGAAGGYTAWTAGTTT  5'-ACARCTTTGAAGGYTAWTAGTTT
Drosophila   ......................     ...G.........C..T......
Mosquito     T..............C......     T..G.........C..C......
Silk moth    .T..T.......A..........     .T.AT.......AT..T......
Asp. beetle  TT.....................     TT.A.........T..A......
Flour beetle ..............G.......     ...A.........T.GT......
Honeybee     TT..T.......A..........     TT.AT.......AT..A......
Dobsonfly    T......................     T..G.........C..A......
Froghopper   TT.....................     TT.A.........T..A......
Kissing bug  TT.....................     TT.A.........T..A......
Locust       .T.....................     .T.A.........T..T......
Stonefly     ......................     ...G.........C..T......
Louse        .T.TA......CTA...AG....     .T.TA......CTA..TAG....
Collembola   TGG............CC......     TGGG.........C.CC......
Daphnia      .A............T.......     .A.G.........C.TT......
Artemia      C..............GG......     C..G.........C.GG......
Human        .GG............C..G..C.     .GGG.........C.CT.G..C.
Turtle       GGG............CTC.G....     GGGG.........CCTC.G....
Chiton       .A.....................     .A.G.........C..T......
Snail        T........-.......CTA...     T..G.....-...C..ACTA...
Earthworm    .A............T.......     .A.A.........T.TT......
```

TW-N1284 is a 23 base primer with two-fold degeneracy at three sites. It works well for most insects, and can be paired with TM-J210 or N2-J586 to amplify all or most of the *nad2* gene. It overlaps TW-J1301 by 17 nucleotides.

```
TW-J1301                    10        20              without degeneracy:
(25mer)   5'-GTT-AAWT-----AAACTAATARCCTTCAAA 5'-GTT-AAWT-----AAACTAATARCCTTCAAA
Drosophila   ...-..C.-----.................    ...-..C.-----.........G........
Mosquito     ...-....-----......G..........    ...-..A.-----......G..G........
Silk moth    ...A....-----.........T.......    ...A..A.-----.........AT.......
Asp. beetle  ...-...G-----......T..........    ...-..AG-----......T..A........
Flour beetle ...A...C-----.......C..........    ...A..TC-----.......C.A........
Honeybee     ...A....-TTAA......T...T.......    ...A..A.-TTAA......T..AT.......
Dobsonfly    ...-...C-----......T..........    ...-..TC-----......T..G........
Froghopper   ...A....-----......T..........    ...A..T.-----......T..A........
Kissing bug  ...-...A-----......T..........    ...-..AA-----......T..A........
Locust       ...-...A-----.................    ...-..TA-----.........A........
Stonefly     ...-....-----.................    ...-..T.-----.........G........
Louse        ...A...------....CT...TAG......    ...A..A------....CT...TAG......
Collembola   ..A-T..C-----......GG..........    ..A-T.TC-----......GG.G........
Daphnia      ...-....-----.......A..........    ...-..T.-----.......A.G........
Artemia      ...-....-----......CC..........    ...-..T.-----......CC.G........
Human        ...-.....---AC.G..C..G..........    ...-..A.---AC.G..C..G.G........
Turtle       .A.C..GCCTATT....C.GAG.........    .A.C..GCCTATT....C.GAGG........
Chiton       ...-...A-----.................    ...-..AA-----.........G........
Snail        ...AT..C-----...TAGT......-....    ...AT.AC-----...TAGT..G...-....
Earthworm    ...-...C-----.......A..........    ...-..AC-----.......A.A........
```

TW-J1301 is a long primer of 25 bases, with two-fold degeneracy at two sites. It spans the DHU loop, which is somewhat variable in length. This length variation has little effect on the utility of this primer. It works for most insects. TW-J1301 overlaps TW-N1284 by 17 nucleotides. Paired with either C1-N1828 or C1-N2353, it amplifies two tRNA genes and the start of *cox1*.

```
TY-N1433                10        20
(28mer)   5'-GGCTG-AATTTTAGGCGATAAATTGTAAA
Drosophila   .....-..G...........G........
Mosquito     .....-..G....................
Silk moth    .....-...AA..A...............
Asp. beetle  .....--..AA-.A..A.....C......
Flour beetle ...CT-G.GA...A...G...........
Honeybee     AT..----AAA..A.A.T...........
Dobsonfly    .....--.A.A.........G........
Froghopper   .....----AAAGA.TA............
Kissing bug  .....A.T.AA......G...........
Locust       .....A.T.A.....T.T..G........
Stonefly     .....AG..A.A...T....G........
Louse        .....--.C..A.A.T.....TC......
Collembola   .....----A.....T.T....C......
Daphnia      ...C.-.GGAG.....AT...........
Artemia      ...C.--...A....T.C...........
Human        .....---AG.G.A..AT.GG.C......
Turtle       .....AGTGG...A...T.GGGC....G.
Chiton       ...C.--..AAA...T.G....C......
Snail        A...T---A..A.A..ATCG.GC......
Earthworm    ..TC.--.GA...AA.AT....C......
```

TY-N1433 is a non-degenerate 28 base primer that is useful for amplifying the end of the *nad2* gene when paired with N2-J586. It could be easily modified in the middle to provide a better match to specific taxa.

```
C1-J1709              10        20       without degeneracy:
(21mer)   5'-AATTGGWGGWTTYGGAAAYTG  5'-AATTGGWGGWTTYGGAAAYTG
Drosophila   ......G..G...........     ......G..G..T.....T..
Mosquito     ......................     ......A..A..T.....T..
Silk moth    ......................     ......A..A..T.....T..
Asp. beetle  ......................     ......T..A..T.....T..
Flour beetle ......................     ......A..A..C.....C..
Sawfly       ...A..................     ...A..A..A..T.....T..
Honeybee     ......................     ......A..A..T.....T..
Dobsonfly    ...C..................     ...C..A..A..T.....T..
Froghopper   ......................     ......T..T..T.....T..
Kissing bug  .........C...........     ......A..C..T.....C..
Locust       ......................     ......A..A..C.....T..
Stonefly     ......................     ......A..A..C.....T..
Louse        ............CT.....     ......A..T..T.CT..T..
Collembola   .............T.....     ......A..A..T..T..T..
Daphnia      ......G..G.....G.....     ......G..G..C..G..C..
Artemia      G.....G........T.....     G.....G..A..T..T..C..
Human        ...C.....C.....C.....     ...C..A..C..T..C..C..
Turtle       ......................     ......T..A..T.....T..
Chiton       ......G..G.....T.....     ......G..G..T..T..C..
Snail        .............C.....     ......T..T..T..C..T..
Earthworm    C.....C..G.....G.....     C.....C..G..T..G..C..
```

C1-J1709 is a 21 base primer that is two-fold degenerate at four sites. It is one of the most reliable primers, and works well with either C1-N2353 or C1-N2776.

```
C1-N1738              10        20          without degeneracy:
(23mer)   5'-TTTATTCGTGGRAATGCYATRTC  5'-TTTATTCGTGGRAATGCYATRTC
Drosophila   ........................     ...........G.....T..G..
Mosquito     ........A.....A........     ........A..G..A..C..A..
Silk moth    ........................     ...........G.....T..A..
Asp. beetle  ....A...A..............     ....A...A..G.....C..A..
Flour beetle ........................     ...........A.....T..A..
Sawfly       ..A.G...A..............     ..A.G...A..A.....C..A..
Honeybee     ........G..............     ........G..G.....T..A..
Dobsonfly    ..............G........     ...........A..G..T..A..
Froghopper   ........................     ...........A.....T..G..
Kissing bug  ........A..............     ........A..G.....T..G..
Locust       ..............A........     ...........A..A..T..A..
Stonefly     ........G.....A........     ........G..G..A..T..A..
Louse        ........A.....C.A......     ........A..G..C.AT..A..
Collembola   .....A........A........     .....A.....G..A..T..A..
Daphnia      ..A.G...A.....G........     ..A.G...A..G..G..C..A..
Artemia      ....AC..G..............     ....AC..G..A.....T..A..
Human        .....G..G.....C........     .....G..G..A..C..C..A..
Turtle       .....A.................     .....A.....A.....T..A..
Chiton       ....AC..G.....A........     ....AC..G..G..A..C..G..
Snail        ..C.....A.....GCT......     ..C.....A..A..GCTC..A..
Earthworm    ..A.GG.................     ..A.GG.....G.....T..A..
```

C1-N1738 is a 23 base primer that is two-fold degenerate at three sites. It can be paired with TW-J1301 to amplify the start of *cox1* in many arthropods.

```
C1-J2195              10        20          without degeneracy:
(23mer)   5'-TGATTCTTTGGWCACCCWGAAGT  5'-TGATTCTTTGGWCACCCWGAAGT
Drosophila   .....T.................     .....T.....T.....T.....
Mosquito     .....T.................     .....T.....A.....T.....
Silk moth    .....T........T........     .....T.....A..T..T.....
Asp. beetle  ..............T........     ...........A..T..T.....
Flour beetle ........C...........G..     ........C..T.....A..G..
Sawfly       ..............T........     ...........A..T..A.....
Honeybee     .....T........T........     .....T.....T..T..A.....
Dobsonfly    ........................     ...........T.....T.....
Froghopper   .....T.....G........G..     .....T.....G.....A..G..
Kissing bug  ........................     ...........A.....T.....
Locust       ........................     ...........T.....A.....
Stonefly     .....T.................     .....T.....T.....T.....
Louse        .....T.................     .....T.....A.....T.....
Collembola   .....T...........C.....     .....T.....T.....C.....
Daphnia      ..G...........T........     ..G........T..T..T.....
Artemia      .....T.....C..T........     .....T.....C..T..T.....
Human        .....T..C..............     .....T..C..T.....T.....
Turtle       ..................C.....     ...........T.....C.....
Chiton       ..G..T........T........     ..G..T.....T..T..T.....
Snail        .....T...........C.....     .....T.....T.....C.....
Earthworm    ........................     ...........T.....A.....
```

C1-J2195 is a 23 base primer that is two-fold degenerate at two sites. It is a slight modification of a primer of the same name in the Simon et al. (1994) list. It works for virtually all animals we have tried, and is useful for bridging the region between *cox1* and *cox2* in arthropods with the ancestral arrangement of these genes.

```
C1-N2353              10        20          without degeneracy:
(23mer)   5'-GCTCGTGTATCAACGTCTATWCC  5'-GCTCGTGTATCAACGTCTATWCC
Drosophila   ........................     ....................T..
Mosquito     .....A.................     .....A..............T..
Silk moth    ............TA........     ............TA.....A..
Asp. beetle  .....A.....T..A........     .....A.....T..A.....T..
Flour beetle .....A.................     .....A..............T..
Sawfly       ...G.A.....T..A........     ...G.A.....T..A.....T..
Honeybee     .....A........A....A...     .....A........A....AT..
Dobsonfly    ..C..G........A........     ..C..G........A.....A..
Froghopper   ..A...........A........     ..A...........A.....A..
Kissing bug  ..A.....G..T...........     ..A.....G..T........A..
Locust       ...........T...........     ...........T........T..
Stonefly     ..............A.....G..     ..............A.....G..
Louse        ..C..A........A..C.....     ..C..A........A..C..A..
Collembola   ...........C..A........     ...........C..A.....T..
Daphnia      ........G.....A..C..G..     ........G.....A..C..G..
Artemia      .....A..G..C.....C.....     .....A..G..C.....C..T..
Human        ........G..T...........     ........G..T........T..
Turtle       .....A..G..T...........     .....A..G..T........T..
Chiton       ..C..A........A..C..C..     ..C..A........A..C..C..
Snail        ..A..C..G.....A.....C..     ..A..C..G.....A.....C..
Earthworm    ..C..G..G..C..A....AG..     ..C..G..G..C..A....AG..
```

C1-N2353 is a 23 base primer with two-fold degeneracy at one site. It is quite reliable when paired with either C1-J1709 or TW-J1301.

```
C1-J2756              10        20      without degeneracy:
(20mer)   5'-ACATTCTTTCCTCARCAYTT  5'-ACATTCTTTCCTCARCAYTT
Drosophila   .....T..C..C........     .....T..C..C..A..T..
Mosquito     ..C..T..C...........     ..C..T..C.....A..T..
Silk moth    .....T.....A........     .....T.....A..A..T..
Asp. beetle  .....T..............     .....T........G..C..
Flour beetle ....................     ..............G..T..
Sawfly       .....T..............     .....T........A..T..
Honeybee     ..T.................     ..T...........A..T..
Dobsonfly    ..C.................     ..C...........A..T..
Froghopper   .....T..............     .....T........A..T..
Kissing bug  ..C.....C...........     ..C.....C.....A..C..
Locust       ....................     ..............A..C..
Stonefly     ........C...........     ........C.....A..T..
Louse        ..T..T..C...........     ..T..T..C.....A..T..
Collembola   G.C........A........     G.C........A..A..T..
Daphnia      ..C..T.....G........     ..C..T.....G..A..T..
Artemia      ..T..T..C...........     ..T..T..C.....A..T..
Human        ..T.....C..A........     ..T.....C..A..A..C..
Turtle       ..C..T..C...........     ..C..T..C.....A..T..
Chiton       ..T..T..............     ..T..T........A..C..
Snail        ..T..T.....A........     ..T..T.....A..A..T..
Earthworm    ...........C........     ...........C..A..C..
```

C1-J2756 is a 20 base primer with two-fold degeneracy at two sites. It works well with C2-N3389 or C2-N3665 to amplify across the region between *cox1* and *cox2*.

```
C1-N2776              10        20        without degeneracy:
(22mer)   5'-GGTAATCAGAGTATCGWCGNGG  5'-GGTAATCAGAGTATCGWCGNGG
Drosophila   .......T..A..A........     .......T..A..A..T..A..
Mosquito     .AA....T..............     .AA....T........T..A..
Silk moth    .A.....T..A...........     .A.....T..A.....T..A..
Asp. beetle  ..........A..A........     ..........A..A..A..A..
Flour beetle ....G..................     ....G...........T..T..
Sawfly       .AA....T..............     .AA....T........T..A..
Honeybee     .A..G..T..A..A........     .A..G..T..A..A..T..T..
Dobsonfly    .......T..............     .......T........T..A..
Froghopper   ......................     ................A..A..
Kissing bug  .......T..............     .......T........T..A..
Locust       .A........A...........     .A........A.....T..T..
Stonefly     ............A........     ............A..T..A..
Louse        .A...CA.ATA..A........     .A...CA.ATA..A..T..A..
Collembola   ............C........     ............C..T..T..
Daphnia      ....G..T.....A..C.....     ....G..T.....A..C..G..
Artemia      .A.....T.C...A........     .A.....T.C...A..A..T..
Human        ....G..C.....A........     ....G..C.....A..T..G..
Turtle       ..........A..A........     ..........A..A..T..T..
Chiton       .A..G.....A..A........     .A..G.....A..A..T..A..
Snail        .......CCTA..C..C.....     .......CCTA..C..C..G..
Earthworm    ....G.....A.....C.....     ....G.....A.....C..A..
```

C1-N2776 is a 22 base primer with four-fold degeneracy at one site, and two-fold degeneracy at another. It is fairly versatile, working in combination with C1-J1709 or C1-J2195 to amplify the central portion of the *cox1* gene in many arthropods.

```
TL2-J3033             10                20
(22mer)   5'-TCT-AATATGGCAGA--------TTAGTGCA
Drosophila   ...-..........--------........
Mosquito     ...-..........--------........
Silk moth    ...-..........-------C.ATA..T.
Asp. beetle  ...-..........--------A.......
Flour beetle ...-.T........--------........
Sawfly       .T.-..........--------A..A....
Honeybee     .T.-..........-------A.A......
Dobsonfly    ...-..........--------.......C
Froghopper   ...-G.........-------AA.......
Kissing bug  ...T...G.......-----AAT.ATA....
Locust       ...-..........--------...A....
Stonefly     ...-..........--------.A......
Louse        ...-..........--------A......G
Collembola   ...-..A........--------.A......
Daphnia      ...-..C........--------AA......
Artemia      .T.-..AG.......AACA----.ATA....
Human        GT.-..G........GCCCGG--.A.TC...
Turtle       AT.-GGGG.......GCCAGGTAA..A....
Earthworm    A.C-..G........--------A......C
```

TL2-J3033 is a non-degenerate 22 base primer that has proven useful for amplifying most or all of the *cox2* gene in many insects. It should be noted that the gap shown is introduced primarily to align the vertebrate sequences, and is not a gap for most invertebrates. It does not work for most Lepidoptera, and should either be modified at the 3' end, or TL2-J3043 used in its place for members of this order.

```
TL2-J3043                        10        20              without degeneracy:
(24mer)   5'-GGCAGA--------TTAGTGYAATGRATTTAA  5'-GGCAGA--------TTAGTGYAATGRATTTAA
Drosophila   ......--------.................     ......--------......C....G......
Mosquito     ......--------.................     ......--------......C....A......
Silk moth    ......-------C.ATA.............     ......-------C.ATA..T....G......
Asp. beetle  ......--------A................     ......--------A.....C....G......
Flour beetle ......--------.............C....     ......--------......C....G.C....
Sawfly       ......--------A..A......A.GC....     ......--------A..A..C...AGGC....
Honeybee     ......-------A.A......T....C....     ......-------A.A....C.T..A.C....
Dobsonfly    ......--------.......C..........     ......--------......CC...G......
Froghopper   ......-------AA.......T.A.......     ......-------AA.....C.T.AA......
Kissing bug  ......-----AAT.ATA..............     ......-----AAT.ATA..C....A......
Locust       ......--------...A....T.A.......     ......--------...A..C.T.AG......
Stonefly     ......--------.A................     ......--------.A....C....G......
Louse        ......--------A......G.CA..C....     ......--------A.....CG.CAA.C....
Collembola   ......--------.A......T...GC....     ......--------.A....C.T..GGC....
Daphnia      ......--------AA......TCA.......     ......--------AA....C.TCAG......
Artemia      ......AACA----.ATA....T.........     ......AACA----.ATA..C.T..G......
Human        ......GCCCGG--.A.TC...TAA..C....     ......GCCCGG--.A.TC.C.TAAA.C....
Turtle       ......GCCAGGTAA..A.....AA.GCC...     ......GCCAGGTAA..A..C..AAGGCC...
Earthworm    ......--------A......C..A.G.....     ......--------A.....CC..AGG.....
```

TL2-J3043 is a 24 base primer with two-fold degeneracy at two sites. It is useful in certain cases where TL2-J3033 fails.

```
C2-N3389                 10        20            without degeneracy:
(23mer)   5'-TACTCATARGATCARTATCAYTG  5'-TACTCATARGATCARTATCAYTG
Drosophila   ..T......CT............     ..T.....ACT...G.....T..
Mosquito     ..T..G...CT............     ..T..G..GCT...A.....T..
Silk moth    ..T.....TCT............     ..T.....TCT...G.....T..
Asp. beetle  ..T.....................     ..T.....A.....A.....T..
Flour beetle ..T.....T..............     ..T.....T.....G.....T..
Sawfly       ..T.....TCT............     ..T.....TCT...A.....T..
Honeybee     ..T.....T..............     ..T.....T.....A.....T..
Dobsonfly    ..T.....T..............     ..T.....T.....G.....T..
Froghopper   ..T.....T..C.....C.....     ..T.....T..C..G..C..T..
Kissing bug  ..T..G...CT............     ..T..G..ACT...A.....T..
Locust       ..T......CT............     ..T.....ACT...G.....T..
Stonefly     ..T......CT............     ..T.....ACT...G.....T..
Louse        ..T.G...T..............     ..T.G...T.....A.....T..
Collembola   ..T.....................     ..T.....A.....A.....T..
Daphnia      ..T......CT............     ..T.....ACT...G.....T..
Artemia      ..T.....................     ..T.....G.....G.....C..
Human        .....G....T......C.....     .....G..G.T...G..C..T..
Turtle       ..T.....T.T............     ..T.....T.T...G.....T..
Chiton       ........CCTC............     ........CCTC..G.....T..
Snail        G....G..TCT............     G....G..TCT...A.....C..
Earthworm    ..T..G..TCT............     ..T..G..TCT...A.....T..
```

C2-N3389 is a 23 base primer two-fold degenerate at three sites. It is a slight modification of the primer of the same name in the Simon et al. (1994) listing, and is useful for amplifying across the *cox1* - *cox2* region when paired with C1-J2756. It overlaps C2-J3396 by 8 residues.

```
C2-J3399              10        20         without degeneracy:
(23mer)   5'-ACAATTGGTCAYCAATGATAYTG  5'-ACAATTGGTCAYCAATGATAYTG
Drosophila   .GT.....................     .GT........T........C..
Mosquito     T..G.A..G..............     T..G.A..G..T........T..
Silk moth    T.T..C..A..............     T.T..C..A..T........C..
Asp. beetle  ......................     ..........T........T..
Flour beetle ........A..............     ........A..C........C..
Sawfly       G.......A..............     G.......A..T........T..
Honeybee     T......................     T.........T........T..
Dobsonfly    G.T....................     G.T........T........C..
Froghopper   T.T..............G.....     T.T........T.....G..C..
Kissing bug  T.T.....C..............     T.T.....C..T........T..
Locust       ........A.GA...........     ........A.GA........C..
Stonefly     ......................     ..........T........C..
Louse        GTC.....AA.............     GTC.....AA.T........T..
Collembola   ...G.A..A..............     ...G.A..A..T........T..
Daphnia      ..TG.A..C..............     ..TG.A..C..T........C..
Artemia      GT..CG..G.....G........     GT..CG..G..T..G.....C..
Human        T.......C........G.....     T.......C..C.....G..C..
Turtle       G.C..A..A..............     G.C..A..A..T........C..
Chiton       GTGG.G..G..............     GTGG.G..G..T........C..
Snail        .G.TA...AAC...G........     .G.TA...AACC..G.....T..
Earthworm    ..T.....G..............     ..T.....G..T........T..
```

C2-J3399 is a 23 base primer with two-fold degeneracy at two sites. It can be paired with TK-N3796 to amplify the end of *cox2*, or with A8-N4061 to amplify across the tRNA-lys/tRNA-asp region, including most of the *atp8* gene. C2-J3399 overlaps C2-N3389 by 11 nucleotides.

```
C2-J3624              10        20     without degeneracy:
(20mer)   5'-ACTCCTGGACGATTAAAYCA  5'-ACTCCTGGACGATTAAAYCA
Drosophila   ....................     ................T..
Mosquito     ..A..A.....T........     ..A..A.....T.....T..
Silk moth    .A...A..T...........     .A...A..T........T..
Asp. beetle  ..A..C......C.G.....     ..A..C......C.G..T..
Flour beetle ..C..C.....CC.C.....     ..C..C.....CC.C..T..
Sawfly       GTA..A......C.T.....     GTA..A......C.T..T..
Honeybee     GT...A......A.T.....     GT...A......A.T..T..
Dobsonfly    ........T...........     ........T........T..
Froghopper   TTA..G..G...........     TTA..G..G........T..
Kissing bug  ..A..A......C.T.....     ..A..A......C.T..C..
Locust       ..A..C......C.......     ..A..C......C....C..
Stonefly     ..A..A......C.......     ..A..A......C....T..
Louse        .A..................     .A...............T..
Collembola   GTA..A.....TC........     GTA..A.....TC....C..
Daphnia      GTA.....T..T........     GTA.....T..T.....T..
Artemia      GAC..A..C..GC.T.....     GAC..A..C..GC.T..T..
Human        .T...C.....TC.......     .T...C.....TC....C..
Turtle       GTC..A........G.....     GTC..A........G..T..
Chiton       GT.........G........     GT.........G.....T..
Snail        GTC..G.....T......AT     GTC..G.....T.....TAT
Earthworm    GTG........TC.......     GTG........TC....C..
```

C2-J3624 is a 20 base primer with two-fold degeneracy at one site. It is not as versatile as C2-J3399 or TK-J3790, but is sometimes useful to amplify the region between the *cox2* and *atp6* genes.

```
C2-N3665              10        20
(20mer)   5'-CCACAAATTTCTGAACATTG
Drosophila   ..G.................
Mosquito     ....................
Silk moth    ....................
Asp. beetle  ....................
Flour beetle ..G.................
Sawfly       ..G.................
Honeybee     ....................
Dobsonfly    ....................
Froghopper   ..................C..
Kissing bug  ..............G.....
Locust       .....G........G.....
Stonefly     .....G..............
Louse        ....................
Collembola   ....................
Daphnia      ...........G..G.....
Artemia      ...........G........
Human        .....G.....A..G.....
Turtle       .....G........G.....
Chiton       ...........A..G..C..
Snail        ..G.....C..C..G.....
Earthworm    ..G..G........G..C..
```

C2-N3665 is a non-degenerate primer that works for almost all Metazoa. It could be made more appropriate for some animals by replacing the ‘T’ with a ‘Y’ at position 18 and replacing the ‘A’ at position 15 with either a ‘G’ or ‘R’. This primer is modified from C2-N-3661 (Simon et al. 1994) by removing 4 nucleotides from the 3' end.

```
TK-J3790              10         20
(23mer)   5'-CATTAGATGACTGAAAGC-AAGTA
Drosophila   ..................-.....
Mosquito     ..................-.....
Silk moth    ..................-.....
Asp. beetle  ....G....G....TTA--...C.
Flour beetle .........G.......T-...C.
Sawfly       .........G.......A-...C.
Honeybee     .........T.....TTAT...AG
Dobsonfly    ..................T-.....
Froghopper   .....AG...........-.....
Kissing bug  .....AG..........T-.....
Locust       ...C.............T-.....
Stonefly     ...C.............-C.....
Collembola   .....A...G.......T-...C.
Daphnia      ...C....AG..T....T-...C.
Artemia      ...AG.G..G.......TC...CG
```

TK-J3790 is a non-degenerate primer that works well with many insects. No reasonable alignment with the louse or Metazoa outside of the Arthropoda was evident, so they are omitted. For many taxa, this primer works well with A6-N4552 to amplify tRNA-asp, *atp8* and most of *atp6*. For taxa that have mismatches near the 3' end, it is necessary to either modify the primer, or simply bridge this region using C2-J3399 paired with A8-N4061.

```
TK-N3796              10        20
(20mer)   5'-ACTATTAGATGGTTTAAGAG
Drosophila   .....A.A............
Mosquito     .....A.T...A........
Silk moth    .....A.A............
Asp. beetle  .....G.TT...A....A.T
Flour beetle .......TG...........
Sawfly       .TC..ATA..AAC.......
Honeybee     TACTA..TT..A.....A..
Dobsonfly    .....A.A............
Froghopper   .....A.T............
Kissing bug  ..C..ATAC...........
Locust       ......TTT...........
Stonefly     .....A.A............
Collembola   .A...A.AG...........
Daphnia      .GC..CTTTAA.C.......
Artemia      ..C..A...GA.C.......
Human        CT..A.CTT.AAC....A..
Turtle       CTCTC.TTC.A.C....A..
Chiton       TA.CG..AC.AAA....A.A
Snail        TTCCG..CG.A.A....A.A
Earthworm    TACTA.CTC.T.A....A..
```

TK-N3796 is a non-degenerate primer that works with many arthropods to amplify all or part of the *cox2* gene when paired with TL2-J3033 or C2-J3399 See the note for TK-J3790 for a alternative strategy for taxa where the primer fails.

```
A8-N4061              10        20             without degeneracy:
(25mer)   5'-GAGAATAAGTTWGTTATCATTTTCA  5'-GAGAATAAGTTWGTTATCATTTTCA
Drosophila   ..A.....A..............     ..A.....A..T............
Mosquito     ........................     ..........T............
Silk moth    ..A.....................     ..A........T............
Asp. beetle  ..A.....A...A...........     ..A.....A..TA...........
Flour beetle .C.....G.....C..........     .C.....G...T.C..........
Sawfly       ..A.........A.........GT     ..A........TA.........GT
Honeybee     TCA.....A...A......AAA...     TCA.....A..TA......AAA...
Dobsonfly    ..A....GA...............     ..A....GA..T............
Froghopper   ..A.....A..............     ..A.....A..T............
Kissing bug  .......G..C............     .......G..CT............
Locust       .......GA...........A....     .......GA..T........A....
Stonefly     ..A..A.G................     ..A..A.G...T............
Louse        ..T.G..GTC..A...AA....AA.     ..T.G..GTC.TA...AA....AA.
Collembola   CTA.....A..........G....     CTA.....A..A........G....
Daphnia      ........A....C..........     ........A..A.C..........
Artemia      ..A..A.G.C.C.C..........     ..A..A.G.C.C.C..........
Human        .C...C.GA...TCGT.......GG     .C...C.GA..TTCGT.......GG
Turtle       TC...G..TG...GGT....GG...     TC...G..TG.T.GGT....GG...
Chiton       ..A..A.TA.C.A...........T     ..A..A.TA.CTA...........T
```

A8-N4061 is a 25 base primer with two-fold degeneracy at one site. It takes advantage of the overlap between the *atp8* and *atp6* genes found in most Metazoan mtDNAs. It is useful with most arthropods for amplifying across the region between the *cox2* and *atp8* genes. This region, with tRNA-lys and tRNA-asp genes, has undergone rearrangement (KD → DK) in a number of taxa, including Hymenoptera and Orthoptera.

```
A6-J4463              10        20          without degeneracy:
(23mer)   5'-TTTGCCCATCTWGTWCCNCAAGG  5'-TTTGCCCATCTWGTWCCNCAAGG
Drosophila   .....T..CT.............     .....T..CT.A..A..T.....
Mosquito     .....T...T.............     .....T...T.A..T..T.....
Silk moth    ...ATT...A..A..........     ...ATT...A.AA.T..A.....
Asp. beetle  ......................     ..........A..A..C.....
Flour beetle ........C..............     ........C..A..T..A.....
Sawfly       ..CA....C............AA     ..CA....C..A..T..T...AA
Honeybee     C..AGT...T........TT.AA     C..AGT...T.A..T..ATT.AA
Dobsonfly    .........T.............     .........T.A..A..T.....
Froghopper   ...TAT......A.....GTG..     ...TAT.....TA.T..TGTG..
Kissing bug  C.C..............G....     C.C........A..T..CG....
Locust       ...A.A........C........     ...A.A.....T..C..A.....
Stonefly     .....T...T.............     .....T...T.A..T..T.....
Louse        ..A......T.GAC....AT...     ..A......T.GACA..AAT...
Collembola   ..A.....C............A.     ..A.....C..A..A..A...A.
Daphnia      C.C......T........TT...     C.C......T.A..T..TTT...
Artemia      ..A..............TT...     ..A........T..A..TTT...
Human        C.A.....CT.CT..........     C.A.....CT.CT.A..A.....
Turtle       C.A.GA......T.....G....     C.A.GA.....AT.A..AG....
Chiton       G.GAGT...T..T.G...AGG..     G.GAGT...T.AT.G..TAGG..
```

A6-J4463 is a 23 base primer with four-fold degeneracy at one site, and two-fold degeneracy at two other sites. It can be paired with either C3-N4908 or C3-N5460 to amplify across the *atp6* - *cox3* junction. Some modification of the 3' end is necessary for Hymenoptera, Hemiptera and a few other groups.

```
A6-N4552              10        20         without degeneracy:
(21mer)   5'-ATGTCCWGCAATYATATTWGC  5'-ATGTCCWGCAATYATATTWGC
Drosophila   ......................     ......A.....T.....A..
Mosquito     ......................     ......T.....T.....A..
Silk moth    ..............G.....     ......T.....T..G..A..
Asp. beetle  ...A...................     ...A..A.....T.....A..
Flour beetle ...G...................     ...G..T.....T.....T..
Sawfly       ......................     ......T.....T.....T..
Honeybee     ........A.....A......     ......A.A...T.A...T..
Dobsonfly    ...C..............G..     ...C..A.....T.....G..
Froghopper   ...A...................     ...A..A.....T.....T..
Kissing bug  ......................     ......T.....T.....A..
Locust       ......................     ......T.....C.....T..
Stonefly     G.....................     G.....A.....T.....T..
Louse        .........T.A.........     ......A..T.AT.....A..
Collembola   ...A...................     ...A..T.....T.....A..
Daphnia      G..G.................     G..G..A.....C.....A..
Artemia      ...G..G..............     ...G..G.....T.....A..
Human        G..G......G.A..G.....     G..G..T...G.A..G..A..
Turtle       ...A.....TG...AG..G..     ...A..T..TG.T.AG..G..
Chiton       .........CC..........     ......A..CC.T.....A..
```

A6-N4552 is a 21 base primer with two-fold degeneracy at three sites. It works well for most insects and can be paired with one of the *cox2* J primers, or TK-J3790 to amplify the *atp8* and most of the *atp6* genes.

```
C3-J4792              10        20     without degeneracy:
(20mer)   5'-GTTGATTATAGACCWTGRCC  5'-GTTGATTATAGACCWTGRCC
Drosophila   ..........C........     ..........C..A..A..
Mosquito     ....................     ............A..A..
Silk moth    ..A.................     ..A..........A..A..
Asp. beetle  .....CC.A..........     .....CC.A.....T..A..
Flour beetle ..A..CGTA..G........     ..A..CGTA..G..A..A..
Sawfly       ..CA..G.A..........     ..CA..G.A.....T..A..
Honeybee     ...ACAA............     ...ACAA.......T..A..
Dobsonfly    ..A........T..C.....     ..A........T..C..A..
Froghopper   ..A...GCA..C........     ..A...GCA..C..T..A..
Kissing bug  ..C..........C.....     ..C..........C..G..
Locust       ..A.................     ..A..........A..A..
Stonefly     ..A...C.A..........     ..A...C.A.....T..A..
Louse        ..A...G.G..G........     ..A...G.G..G..T..A..
Collembola   ..A..CC.A.....G.....     ..A..CC.A.....G..A..
Daphnia      ..A...A.A..C..C.....     ..A...A.A..C..C..G..
Artemia      ...A..AT............     ...A..AT......T..A..
Human        ..AA.ACCC..C........     ..AA.ACCC..C..A..A..
Turtle       ..A..CCCA..C..G.....     ..A..CCCA..C..G..A..
Chiton       ..A..A.T...T........     ..A..A.T...T..T..G..
Snail        .....AATA..C........     .....AATA..C..T..A..
Earthworm    .....G..C..C........     .....G..C..C..A..G..
```

C3-J4792 is a 20 base primer with two-fold degeneracy at two sites. It is fairly versatile, and works with a wide range of arthropods when paired with C3-N5460 or N3-N5731. It is based on a region that is somewhat variable at the 3' end, and may have to be modified for specific taxa. In that case, it is necessary to sequence this region to identify the appropriate modifications. See primer A6-J4463 for a strategy.

```
C3-J4908              10        20         without degeneracy:
(23mer)   5'-CGAGTTAYATCTCGTCATCATTG  5'-CGAGTTAYATCTCGTCATCATTG
Drosophila   ..T.AA..................     ..T.AA.C...............
Mosquito     ....AA..................     ....AA.T...............
Silk moth    ..T.AA..................     ..T.AA.T...............
Asp. beetle  ..T..................AA.     ..T....C............AA.
Flour beetle ....AA..G..............     ....AA.TG..............
Sawfly       ...A....................     ...A...C...............
Honeybee     ...A.A........AA.....AA     ...A.A.T......AA.....AA
Dobsonfly    ..C..G..................     ..C..G.T...............
Froghopper   ..T........A...........     ..T....C...A...........
Kissing bug  ..G..G..................     ..G..G.T...............
Locust       ...AC............C.....     ...AC..C.........C.....
Stonefly     ..T..G.....A...........     ..T..G.T...A...........
Louse        GA.A...........AT...CAT     GA.A...T.......AT...CAT
Collembola   ..T.A....C..........C..     ..T.A..C.C..........C..
Daphnia      ..GAC...G........C.....     ..GAC..CG........C.....
Artemia      ..T.AC..G..............     ..T.AC.CG..............
Human        ..T........G..C........     ..T....C...G..C........
Turtle       ..TACG.....C...........     ..TACG.C...C...........
Chiton       ...A..........C..C.....     ...A...C......C..C.....
Snail        .....A..G...........CAA     .....A.CG...........CAA
Earthworm    ..CACG........C.....C..     ..CACG.C......C.....C..
```

C3-N4908 is a 23 base primer with two-fold degeneracy at one site. It is fairly versatile, but some modifications are needed at the 3' end for certain taxa.

```
C3-N5460              10        20     without degeneracy:
(20mer)   5'-TCAACAAAATGTCARTAYCA  5'-TCAACAAAATGTCARTAYCA
Drosophila   ....................     ..............G..T..
Mosquito     ....................     ..............A..T..
Silk moth    ..T.................     ..T...........A..T..
Asp. beetle  ....................     ..............G..T..
Flour beetle ....................     ..............A..T..
Sawfly       ..T.....G...........     ..T.....G.....G..T..
Honeybee     ..T.................     ..T...........A..T..
Dobsonfly    ....................     ..............G..T..
Froghopper   ....................     ..............G..T..
Kissing bug  ....................     ..............G..T..
Locust       ........G...........     ........G.....G..T..
Stonefly     ........G...........     ........G.....A..T..
Louse        ....................     ..............A..T..
Collembola   ..T.................     ..T...........G..T..
Daphnia      ........G..C........     ........G..C..A..T..
Artemia      ....................     ..............A..T..
Human        ..T........C........     ..T........C..G..T..
Turtle       ..T..G..G...........     ..T..G..G.....A..T..
Chiton       ...........C........     ...........C..G..C..
Snail        ...CA..GG...........     ...CA..GG.....A..T..
Earthworm    ..T.....G...........     ..T.....G.....G..T..
```

C3-N5460 is a 20 base primer with two-fold degeneracy at two sites. It is highly conserved and should work with virtually all Metazoa. It overlaps C3-J5470 by 11 residues. It is a slight modification of the primer of the same name in Simon et al. (1994).

```
C3-J5470              10        20     without degeneracy:
(20mer)   5'-GCAGCTGCYTGATAYTGRCA  5'-GCAGCTGCYTGATAYTGRCA
Drosophila   ........A...........     ........A.....C..A..
Mosquito     ....................     ........C.....T..A..
Silk moth    ..T..A..............     ..T..A..T.....T..A..
Asp. beetle  .....A..............     .....A..T.....C..A..
Flour beetle ..T..A..............     ..T..A..C.....T..A..
Sawfly       ........A...........     ........A.....C..A..
Honeybee     TT...AAT............     TT...AATT.....T..A..
Dobsonfly    ....................     ........T.....C..A..
Froghopper   ....................     ........T.....C..A..
Kissing bug  ........A...........     ........A.....C..A..
Locust       ..T..A..A...........     ..T..A..A.....C..A..
Stonefly     ....................     ........T.....T..A..
Louse        .....A..............     .....A..T.....T..A..
Collembola   ..T..A..............     ..T..A..C.....C..A..
Daphnia      ..G..A..A...........     ..G..A..A.....T..G..
Artemia      ..T..A..............     ..T..A..T.....T..A..
Human        ..C..C..............     ..C..C..C.....C..G..
Turtle       ....................     ........C.....T..A..
Chiton       ...........G........     ........T..G..C..G..
Snail        ....................     ........T.....T..A..
Earthworm    ..C.................     ..C.....C.....C..A..
```

C3-J5470 is a 20 base primer with two-fold degeneracy at three sites. It works for most arthropods, and with a little modification, should work with most Metazoa. It overlaps C3-N5460 by 11 residues.

```
N3-N5731              10        20     without degeneracy:
(20mer)   5'-TTAGGGTCAAATCCRCAYTC  5'-TTAGGGTCAAATCCRCAYTC
Drosophila   ....................     ..............A..T..
Mosquito     A.T........C........     A.T........C..G..T..
Silk moth    ..T..A.....C........     ..T..A.....C..A..T..
Asp. beetle  ..G........C........     ..G........C..A..T..
Flour beetle ....................     ..............G..T..
Sawfly       ..G........A........     ..G........A..A..T..
Honeybee     A.T..A.T............     A.T..A.T......A..T..
Dobsonfly    A.T..A.....C........     A.T..A.....C..A..T..
Froghopper   A.T..A..............     A.T..A........A..T..
Kissing bug  ........G...........     ........G.....A..T..
Locust       ....................     ..............A..T..
Stonefly     ....................     ..............A..T..
Louse        .CT.CTCG...A........     .CT.CTCG...A..A..T..
Collembola   AGT....TG...........     AGT....TG.....A..T..
Daphnia      ..T.C......C........     ..T.C......C..A..T..
Artemia      AA...A..G..C........     AA...A..G..C..G..T..
Human        A.......G..G........     A.......G..G..G..C..
Turtle       AAT..A..............     AAT..A........A..T..
Chiton       .....A.....C........     .....A.....C..A..C..
Snail        AGTT....G.C..T..TACT     AGTT....G.C..TA.TACT
Earthworm    ..T........C........     ..T........C..A..C..
```

N3-N5731 is a 20 base primer with two-fold degeneracy at two sites. It is based on the only highly conserved region in the *nad3* gene, and works for most taxa where it has been tried. It overlaps N3-J5747 by 17 bases.

```
N3-J5747              10        20          without degeneracy:
(23mer)   5'-CCATTTGAATGTGGRTTTGAYCC  5'-CCATTTGAATGTGGRTTTGAYCC
Drosophila   ..T.....................     ..T...........A.....C..
Mosquito     ..T........C............     ..T........C..G.....C..
Silk moth    ........................     ..............G.....T..
Asp. beetle  ..C.....................     ..C...........G.....C..
Flour beetle ..........C............     ...........C..A.....C..
Sawfly       ..............T........     ..............T.....C..
Honeybee     ..................A....     ..............A...A.T..
Dobsonfly    ........................     ..............G.....T..
Froghopper   ........................     ..............A.....T..
Kissing bug  ................C.....     ..............A..C..C..
Locust       ........................     ..............A.....C..
Stonefly     ..T.....................     ..T...........A.....C..
Louse        GAG...........T...CGAG.     GAG...........T...CGAG.
Collembola   ................CA....     ..............A..CA.C..
Daphnia      ..T................G.     ..T...........G.....CG.
Artemia      ..........C.....C.....     ...........C..G..C..T..
Human        ..T.AC..G..C..C..C.....     ..T.AC..G..C..C..C..C..
Turtle       ....AC..................     ....AC........A.....T..
Chiton       ..T.....G...............     ..T.....G.....G.....T..
Snail        TT..CAAGTA..A..G.C....A     TT..CAAGTA..A.AG.C..C.A
Earthworm    .....C..G...............     .....C..G.....G.....C..
```

N3-J5747 is a 23 base primer with two-fold degeneracy at two sites. It works well for most arthropods and should work for many other Metazoa, as well. It overlaps N3-N5731 by 17 nucleotides.

```
TN-J6155              10         20         without degeneracy:
(22mer)       TTTAATTGAARCCAAAA-AGAGG  5'-TTTAATTGAARCCAAAA-AGAGG
Drosophila    .................-.....     ..........G......-.....
Mosquito      .................-.....     ..........A......-.....
Silk moth     A...............TC.....     A.........A.....TC.....
Asp. beetle   .................-.....     ..........A......-.....
Flour beetle  .................-.....     ..........A......-.....
Sawfly        .......A.........T.....     .......A..G......T.....
Honeybee      ....G..A....T...TA.TTA.     ....G..A..G.T...TA.TTA.
Dobsonfly     .................-.....     ..........A......-.....
Froghopper    .....C.A.........T.....     .....C.A..G......T.....
Kissing bug   A.................-.....     A.........G......-.....
Locust        A...............G-.....     A.........G.....G-.....
Stonefly      A...............T-.....     A.........G.....T-.....
Louse         .......A.........-.....     .......A..G......-.....
```

TN-J6155 is a 22 base primer with two-fold degeneracy at one site. It is well conserved across insects, but no reasonable alignment with other arthropods, or other Metazoa is evident. For many insects, this primer, or TN-J6172 (below), can be paired with N5-N7211 to amplify the end of the *nad5* gene and tRNAs for Ser, Glu and Phe.

```
TN-N6160              10         20         without degeneracy:
(22mer)    5'-TCAATTTTRTCATTAACAGTGA  5'-TCAATTTTRTCATTAACAGTGA
Drosophila    ......A..............     ......A.A.............
Mosquito      ....................     ........A.............
Silk moth     ....................     ........A.............
Asp. beetle   ....................     ........G.............
Flour beetle  ....................     ........G.............
Sawfly        ...T.AAA.........A...     ...T.AAAA.........A...
Honeybee      ...TA.GAT.T.......A.A.     ...TA.GAT.T.......A.A.
Dobsonfly     .....AG..............     .....AG.A.............
Froghopper    ........T.T.......A.A.     ........T.T.......A.A.
Kissing bug   ...T....T.........A...     ...T....T.........A...
Locust        ...T..A...T.........A.     ...T..A.A.T.........A.
Stonefly      ......A..............     ......A.A.............
Louse         ...G..CC..AG.......CT.     ...G..CCA.AG.......CT.
Collembola    G.CT.AACT............     G.CT.AACT............
Daphnia       C.CTAAA.CC...........     C.CTAAA.CC...........
Artemia       C.TCC..CT.........A...     C.TCC..CT.........A...
```

TN-N6160 is a 22 base primer with two-fold degeneracy at one site. It is fairly versatile for arthropods, but no reasonable alignment with other Metazoa is evident. Therefore they are omitted from the alignment. This tRNA is third in a set of six (ARNSEF) between the *nad3* and *nad5* genes in the ancestral arthropod arrangement. It is particularly useful for bridging this region, and can be paired with C3-J5488 or N3-N5747. It overlaps TN-J6172 by 13 bases.

```
TN-J6172              10        20
(22mer)   5'-AGAGGTATATCACTGTTAATGA
Drosophila   .....CG..............
Mosquito     ......................
Silk moth    ......................
Asp. beetle  ........G.............
Flour beetle ........G.............
Sawfly       .....C.AG...T.........
Honeybee     .TTA.C.AT.T.T.......A.
Dobsonfly    ......................
Froghopper   .....C.AT.T.T.......A.
Kissing bug  .....CC.T...T.........
Locust       .....CG...T.........A.
Stonefly     .....C...............
Louse        .....C.C..AG.......CT.
Collembola   TATA..GCT.............
Daphnia      ..CA.C..C............G
Artemia      GATA.CGCT...T.........
```

TN-J6172 is a non-degenerate 22 base primer. It overlaps TN-N6160 by 13 bases. It works for a wide range of arthropods, but may require modification for some taxa. Paired with N5-N7211, it allows amplification of the difficult region that includes the 3' end of the *nad5* gene in taxa with the ancestral arthropod gene arrangement. The match with Metazoa outside of arthropods is poor, so these are omitted.

```
TF-N6384              10        20           without degeneracy:
(24mer)   5'-TATATTTAGAGYATRAYAYTGAAG  5'-TATATTTAGAGYATRAYAYTGAAG
Drosophila   ........................     ...........T..G.C.C.....
Mosquito     ........................     ...........C..G.C.C.....
Silk moth    ATAT.A......T...........     ATAT.A.....TT.A.T.T.....
Asp. beetle  ..A...A........G........     ..A...A....C..AGT.T.....
Flour beetle ....A...................     ....A......C..G.T.T.....
Sawfly       ..AG....A...T.TT........     ..AG....A..CT.TTC.T.....
Honeybee     ............G.........A     ...........CG.A.T.T....A
Dobsonfly    ..--A.A.................     ..--A.A....C..A.C.T.....
Froghopper   A.A...A.A..............     A.A...A.A..C..A.T.T.....
Kissing bug  .TA.....A....A.G........     .TA.....A..C.AAGT.T.....
Locust       AT.G.G..................     AT.G.G.....T..A.C.T.....
Stonefly     ....AA..................     ....AA.....C..G.C.C.....
Louse        CT..AA..A...G..G.T......     CT..AA..A..CG.AGTTT.....
Collembola   .TAT.AC........G........     .TAT.AC....C..AGT.T.....
Daphnia      A.AT.AA.A......G........     A.AT.AA.A..C..AGC.C.....
Artemia      CT..A.A......C..........     CT..A.A....T.CA.C.C.....
Human        CC.CC.C.A....AT........A     CC.CC.C.A..C.AT.C.C....A
Turtle       .TATCAC.A....C.G........     .TATCAC.A..C.CGGC.C.....
Chiton       .TATA.C......CTG........     .TATA.C....C.CTGC.T.....
Snail        CT...A..A....CCG.T......     CT...A..A..C.CCGCTT.....
Earthworm    .TA.AA..A......T........     .TA.AA..A..C..GTT.T.....
```

TF-N6384 is a 24 base primer with two-fold degeneracy at four sites. It works for a wide range of taxa, in many arthropod orders, but the region is subject to variation in some taxa. It can be paired with C3-J5488 or N3-J5747 to amplify most of the tRNAs in this region. The position and orientation of this gene is highly conserved. This primer overlaps TF-J-6400 by 17 nucleotides.

```
TF-J6400              10        20         without degeneracy:
(23mer)   5'-TAACATCTTCAATRTYATRCTCT  5'-TAACATCTTCAATRTYATRCTCT
Drosophila   ..........G..........     ..........G.G.C..A....
Mosquito     ..........G..........     ..........G.G.C..G....
Silk moth    ...T............A.....     ...T.........A.T.AA....
Asp. beetle  ..GT.G........C........     ..GT.G.......ACT..G....
Flour beetle .G.T...................     .G.T.........A.C..G....
Sawfly       .TT..........AA.A...T.     .TT..........GAA.AG..T.
Honeybee     ...T..T.........C.....     ...T..T......A.T.CG....
Dobsonfly    .....G..............     .....G.......G.T..G....
Froghopper   ...T................T.     ...T.........A.T..G..T.
Kissing bug  ..GT.........C.T....T.     ..GT.........ACTT.G..T.
Locust       ....................     ............G.T..A....
Stonefly     ..........G..........     ..........G.G.C..G....
Louse        ..GT........A.C..C...T.     ..GT........AACT.CG..T.
Collembola   ..GT.........C........     ..GT.........ACT..G....
Daphnia      ..G........G..C......T.     ..G........G.GCT..G..T.
Artemia      C.........G....G......     C..........G.G.TG.A....
Human        A.....T....G...AT....T.     A.....T....G.G.AT.G..T.
Turtle       .GG..A.....G..C.G....T.     .GG..A.....G.GCCG.G..T.
Chiton       .TG..G........CAG......     .TG..G.......GCAG.G....
Snail        GCG.C.......A.CGG....T.     GCG.C.......AGCGG.G..T.
Earthworm    .GCT.........A......T.     .GCT.........AAC..G..T.
```

TF-J6400 is a 23 base primer with two-fold degeneracy at three sites. It could be made more versatile by introducing degeneracy at the 3' end, replacing the 'C' with a 'Y'. It overlaps TF-N6384 by 17 bases.

```
N5-J7077              10        20         without degeneracy:
(22mer)   5'-TTAAATCCTTWGARTAAAAYCC  5'-TTAAATCCTTWGARTAAAAYCC
Drosophila   ...T................     ...T......A..A.....T..
Mosquito     .C..................     .C........T..A.....C..
Silk moth    .......T............     .......T..A..A.....T..
Asp. beetle  ....G..T............     ....G..T..T..G.....T..
Flour beetle CC..................     CC........T..A.....T..
Sawfly       ....................     ..........A..A.....C..
Honeybee     .......T.........T.A..     .......T..T..A...T.A..
Dobsonfly    ....................     ..........A..A.....T..
Froghopper   ....................     ..........T..G.....T..
Kissing bug  .A.TG...............     .A.TG.....A..A.....T..
Locust       ....................     ..........A..A.....C..
Stonefly     ....................     ..........T..G.....T..
Louse        .......T............     .......T..T..A.....C..
Collembola   ...T....C............     ...T....C.A..A.....T..
Daphnia      ....................     ..........A..G.....C..
Artemia      ...GG..T......A.G..A..     ...GG..T..T..AA.G..A..
Human        .GTGG..T..G.....G..A..     .GTGG..T..G..G..G..A..
Turtle       .CGC...T..G......T....     .CGC...T..G..G...T.T..
Chiton       ...C...T............     ...C...T..A..G.....T..
Snail        .AGC..G.........GT.AG.     .AGC..G...T..A..GT.AG.
Earthworm    ...T..............A..     ...T......A..G.....A..
```

N5-J7077 is a 22 base primer with two-fold degeneracy at two sites. It works fairly consistently with most arthropods when paired with N5-N7793, amplifying a major portion of the *nad5* gene.

```
N5-N7211              10        20             without degeneracy:
(25mer)   5'-TTAAGGCTTTAYTATTTATRTGYGC  5'-TTAAGGCTTTAYTATTTATRTGYGC
Drosophila   ....A.....................     ....A......T.......A..T..
Mosquito     .......G.................     .......G...T.......A..T..
Silk moth    ....A.....................     ....A......T.......A..T..
Asp. beetle  ....A.....................     ....A......T.......A..T..
Flour beetle .C........G..G............     .C........GT.G.....A..T..
Sawfly       ....AT...........T.....T.     ....AT.....T.....T.A..TT.
Honeybee     ....AT.A...A.G..........T     ....AT.A...A.G.....A..T.T
Dobsonfly    ........................     ...........T.......A..T..
Froghopper   ....A............T.......     ....A......T.....T.A..T..
Kissing bug  .............T...T.......     ...........C.T...T.G..T..
Locust       ................C........     ...........T....C..G..T..
Stonefly     ........................     ...........T.......A..T..
Louse        ....A..AGC.ACT......AT...     ....A..AGC.ACT.....AATT..
Collembola   ....AT....G......T.......     ....AT....GT.....T.A..T..
Daphnia      .......CC.G..T...C.T.....     .......CC.GC.T...C.T..T..
Artemia      ........C.T..T.........T.     ........C.TC.T.....A..TT.
Human        .C..A..CA..............T.     .C..A..CA..C.......G..CT.
Turtle       ....A..CA...............     ....A..CA..T.......A..C..
Chiton       ....A..CC.C.....CC.T.....     ....A..CC.CC....CC.T..C..
Snail        ....A...C.T..G...C..GC.C.     ....A...C.TT.G...C.AGCCC.
Earthworm    ........C.T..C...G.......     ........C.TC.C...G.G..C..
```

N5-N7211 is a 25 base primer with two-fold degeneracy at three sites. It works well for a wide range of arthropods and can be paired with TN-J6155, TN-J6172 or TF-J6400 to amplify the end of the *nad5* gene.

```
N5-J7572              10        20           without degeneracy:
(24mer)   5'-AAAGGGAATTTGAGCTCTTTTWGT  5'-AAAGGGAATTTGAGCTCTTTTWGT
Drosophila   ...A......................     ...A..................A..
Mosquito     ...A...........A........     ...A...........A.....T..
Silk moth    ...A...........A........     ...A...........A.....A..
Asp. beetle  ...A............GA......     ...A............GA...T..
Flour beetle ...C......................     ...C..................T..
Sawfly       ...A.....C..............     ...A.....C............A..
Honeybee     ...A........T...........     ...A........T........A..
Dobsonfly    ...A......................     ...A..................A..
Froghopper   ......T.A.........C......     ......T.A.........C..A..
Kissing bug  ...C.....C........CC....     ...C.....C........CC.A..
Locust       ...A........T.....C.....     ...A........T.....C..A..
Stonefly     ...A..............C.....     ...A..............C..A..
Louse        ...T....A...............     ...T....A............T..
Collembola   ...A..............CC....     ...A..............CC.A..
Daphnia      G.......................     G...................A..
Artemia      ...A..T.G............G..     ...A..T.G............G..
Human        G.GACCT.A...G...GA...GCC     G.GACCT.A...G...GA...GCC
Turtle       ..GTCCG.A.......GA....CC     ..GTCCG.A.......GA...TCC
Chiton       ...T..T.G...............     ...T..T.G............A..
Snail        ...T...TA...T...........     ...T...TA...T........T..
Earthworm    ...C..C..C..G...........     ...C..C..C..G........A..
```

N5-J7572 is a 24 base primer with two-fold degeneracy at one site. This primer is quite reliable with most insects and should work with other arthropods, as well. It can be paired with N4-N8727 to produce a 1.2 kb fragment connecting the genes *nad5* with *nad4*.

```
N5-N7793              10        20     without degeneracy:
(20mer)   5'-TTAGGTTGRGATGGNYTAGG  5'-TTAGGTTGRGATGGNYTAGG
Drosophila   .....G.............     .....G..A.....AT....
Mosquito     C.T..A.............     C.T..A..A.....GT....
Silk moth    ..G................     ..G.....G.....TT....
Asp. beetle  ................C..     ........G.....GC.C..
Flour beetle .....G.............     .....G..G.....GC....
Sawfly       ..G..G.............     ..G..G..A.....GT....
Honeybee     ...................     ........A.....TT....
Dobsonfly    ..G..A.............     ..G..A..A.....GC....
Froghopper   ...................     ........A.....TT....
Kissing bug  A.T................     A.T.....G.....TT....
Locust       ...................     ........A.....AT....
Stonefly     .....G.............     .....G..A.....TT....
Louse        G....A......AT......     G....A..A...ATGT....
Collembola   ...................     ........G.....GT....
Daphnia      C....A..........G..     C....A..G.....GT.G..
Artemia      ...................     ........A.....GC....
Human        A.C..C.....G...G....     A.C..C..A..G..CG....
Turtle       A.T........A...G....     A.T.....G..A..AG....
Chiton       C....C..........C..     C....C..A.....GC.C..
Snail        A....C..........G.T     A....C..A.....CT.G.T
Earthworm    C.T........C.....G..     C.T.....A..C..CT.G..
```

N5-N7793 is a 20 base primer with four-fold degeneracy at one site, and two-fold degeneracy at two other sites. It works for many insects when paired with N5-J7077.

```
N5-J7806              10        20          without degeneracy:
(23mer)   5'-GAMACAARACCTAACCCATCYCA  5'-GAMACAARACCTAACCCATCYCA
Drosophila   ........T.....T........     ..A....GT.....T.....T..
Mosquito     .....T..T..............     ..A..T.AT...........T..
Silk moth    ....TT........A........     ..A.TT.A......A.....C..
Asp. beetle  ...........G.G.........     ..A....G...G.G......C..
Flour beetle .....T.......G.........     ..A..T.A.....G......C..
Sawfly       ..T..T.................     ..T..T.A............T..
Honeybee     ....TT..T.....A........     ..A.TT.AT.....A.....T..
Dobsonfly    .....T..T....G.........     ..A..T.GT....G......T..
Froghopper   ....TC........A........     ..A.TC.A......A.....T..
Kissing bug  ..............A........     ..C....G......A.....C..
Locust       ..T..T........T........     ..T..T.A......T.....T..
Stonefly     ..T..T..T.....A........     ..T..T.AT.....A.....T..
Louse        ...GAT.C.......AT......     ..AGAT.C.......AT...T..
Collembola   .....C..T..............     ..A..C.AT...........C..
Daphnia      ....TT..C..C...........     ..A.TT.GC..C........C..
Artemia      ...GTG.T.....G.........     ..AGTG.T.....G......T..
Human        ..T.T..TT....CG..C.....     ..T.T..TT....CG..C..T..
Turtle       ..T.TG.CT....CT..T.....     ..T.TG.CT....CT..T..C..
Chiton       ..T..T..C..G.G.........     ..T..T.GC..G.G......T..
Snail        CG.CG...TA.C..G........     CGACG..ATA.C..G.....T..
Earthworm    ..T....TT..C..G..G.....     ..T....TT..C..G..G..T..
```

N5-J7806 is a 23 base primer with two-fold degeneracy at three sites. It works for a wide range of insects, producing a 1 kb fragment when paired with N4-N8727.

```
N4-N8487              10        20        without degeneracy:
(22mer)   5'-TATCAGSTAATATRGCWGCTCC  5'-TATCAGSTAATATRGCWGCTCC
Drosophila   G.....................     G.....C......A..A.....
Mosquito     GT.....C..............     GT....GC.....A..A.....
Silk moth    ....CT.A..C...........     ....CTCA..C..A..A.....
Asp. beetle  ..ATT.................     ..ATT.G......G..A.....
Flour beetle ..G..A.A.....T........     ..G..AGA.....T..T.....
Sawfly       GT..TT........T.G.....     GT..TTC......AT.G.....
Honeybee     GT...T.A.......G.T....     GT...TCA.....A.GAT....
Dobsonfly    .T...A.A...........C..     .T...AGA.....G..T..C..
Froghopper   GT..TT.A.........AGA..     GT..TTCA.....A..TAGA..
Kissing bug  G...TAA..........T....     G...TAA......G..TT....
Locust       G...TT..........T....     G...TTC......G..TT....
Stonefly     GT...T..........G..A..     GT...TG......A..G..A..
Louse        ...GTATA...TC..G.ATC..     ...GTATA...TCA.GAATC..
Collembola   GTGT.........TT.......     GTGT..C......TT.A.....
Daphnia      CTATTA.............C..     CTATTAG......A..T..C..
Artemia      GC.TTCTC.......G.TTA..     GC.TTCTC.....A.GATTA..
Human        GCCTC.....CC.C..CTTA..     GCCTC.C...CC.C..CTTA..
Turtle       GC.T...C..C.....CAT...     GC.T..CC..C..A..CAT...
Chiton       GT...ATC..C...........     GT...ATC..C..A..A.....
Snail        GCAGGT.A.TA...........     GCAGGTGA.TA..A..A.....
Earthworm    C.G.TA.A...........C..     C.G.TAGA.....A..A..C..
```

N4-N8487 is a 22 base primer with two-fold degeneracy at three sites. It works well for some taxa, but not for others. Taxon-specific modifications, particularly at the 3' end would make it particularly useful for examining the *nad5* - *nad4* junction.

```
N4-J8641              10        20     without degeneracy:
(20mer)   5'-CCAGAAGAACATAANCCRTG  5'-CCAGAAGAACATAANCCRTG
Drosophila   ....................     ..............A..A..
Mosquito     ....................     ..............C..A..
Silk moth    ....................     ..............A..A..
Asp. beetle  .....T..G...........     .....T..G.....A..A..
Flour beetle ....................     ..............C..A..
Sawfly       ....................     ..............A..A..
Honeybee     ..T.......T.........     ..T.......T...A..A..
Dobsonfly    ....................     ..............T..A..
Froghopper   ............G......     ............GC..A..
Kissing bug  ...........C........     ...........C..A..A..
Locust       ..........TA.T......     ..........TA.TT..A..
Stonefly     T.C........A.G......     T.C........A.GA..A..
Louse        AA.......GTA........     AA.......GTA..C..A..
Collembola   .........G.A.G......     .........G.A.GC..A..
Daphnia      ..C........G.G......     ..C........G.GG..G..
Artemia      .....G.....G.T......     .....G.....G.TC..G..
Human        AAT..G..TGTA.G......     AAT..G..TGTA.GC..G..
Turtle       ATT..G..TGT..G......     ATT..G..TGT..GG..A..
Chiton       ..T.....GG.A.G......     ..T.....GG.A.GA..G..
Snail        ..T..C..TGCCC.......     ..T..C..TGCCC.A..G..
Earthworm    G.G..T..G.TC.G......     G.G..T..G.TC.GC..A..
```

N4-J8641 is a 20 base primer with four-fold degeneracy at one site, and two-fold degeneracy at another site. It works well for many insects, and can be paired with N4-N9153 or N4L-N9629.

```
N4-N8727              10        20          without degeneracy:
(23mer)   5'-AAATCTTTRATTGCTTATTCWTC  5'-AAATCTTTRATTGCTTATTCWTC
Drosophila   ..GG.........A.........     ..GG....A.....A.....A..
Mosquito     ..GG...................     ..GG....A...........T..
Silk moth    .......................     ........A...........T..
Asp. beetle  .......................     ........A...........A..
Flour beetle .......................     ........A...........T..
Sawfly       ..G..A.................     ..G..A..A...........A..
Honeybee     .....AA.T.....AAT......     .....AA.T.....AAT...A..
Dobsonfly    .....AA..............AG     .....AA.G...........TAG
Froghopper   ..............C........     ........A.....C.....T..
Kissing bug  .....A.................     .....A..G...........T..
Locust       ..G....................     ..G.....A...........T..
Stonefly     ..GG...................     ..GG....A...........T..
Louse        ...AGGA.T..............     ...AGGA.T...........T..
Collembola   ...G.....G.A...........     ...G....AG.A........A..
Daphnia      .....G...G.A.....C.....     .....G..AG.A.....C..T..
Artemia      .....C..............C..     .....C..A...........C..
Human        .....GC.C.....A..C.....     .....GC.C.....A..C..T..
Turtle       .....A........C..C.....     .....A..G.....C..C..A..
Chiton       ...G.A.....C........C..     ...G.A..A..C........C..
Snail        ...G.A........C...AGGAG     ...G.A..A.....C...AGGAG
Earthworm    .....CC..........C.....     .....CC.A........C..A..
```

N4-N8727 is a 23 base primer with two-fold degeneracy at two sites. It works well with a very wide range of insects, and has a good match to many other Metazoa. Paired with either N5-J7572 or N5-J7806, it can be used to bridge the difficult region between *nad5* and *nad4* in the typical arthropod gene arrangement.

```
N4-J8941              10        20          without degeneracy:
(23mer)   5'-GAAACAGGAGCCTCAACATGWGC  5'-GAAACAGGAGCCTCAACATGWGC
Drosophila   ..T........T...........     ..T........T........A..
Mosquito     .....T.....T...........     .....T.....T........A..
Silk moth    .....T..G..T...........     .....T..G..T........A..
Asp. beetle  .....C.....T...........     .....C.....T........A..
Flour beetle ..C..............G.....     ..C..............G..A..
Sawfly       ....T...G.....T........     ....T...G.....T.....A..
Honeybee     T..TA......T...........     T..TA......T........A..
Dobsonfly    .C..T...G..T..C........     .C..T...G..T..C.....A..
Froghopper   ....T......T..T........     ....T......T..T.....A..
Kissing bug  .C..T...C..............     .C..T...C...........A..
Locust       ....TT.....T...........     ....TT.....T........A..
Stonefly     ..G........T...........     ..G........T........A..
Louse        ATC.AT..T..T..T........     ATC.AT..T..T..T.....A..
Collembola   CCC..T........C........     CCC..T........C.....T..
Daphnia      .C...T..G...........G..     .C...T..G...........G..
Artemia      .TT..T........T..G.....     .TT..T........T..G..A..
Human        .CG.TG..G..T..G.....G..     .CG.TG..G..T..G.....G..
Turtle       .CG.TT..G..T..T..G.....     .CG.TT..G..T..T..G..T..
Chiton       .CT.....G..T..T........     .CT.....G..T..T.....A..
Snail        .CT.A.T.GCG.CTC..CATGTG     .CT.A.T.GCG.CTC..CATGTG
Earthworm    .CG.TT..G..T..T..G.....     .CG.TT..G..T..T..G..T..
```

N4-J8941 is a 23 base primer with two-fold degeneracy at one site. This primer is modified from the region covered by the primer N4-J-8944 from Simon et al. (1994), making it more versatile.

```
N4-N9153              10        20     without degeneracy:
(20mer)   5'-TGAGGTTATCAACCNGARCG  5'-TGAGGTTATCAACCNGARCG
Drosophila   ....................     ..............G..A..
Mosquito     .....G.....G........     .....G.....G..T..G..
Silk moth    ....................     ..............A..G..
Asp. beetle  ..G..C.....G........     ..G..C.....G..T..A..
Flour beetle .....A..C...........     .....A..C.....G..G..
Sawfly       ..G..G......T....T..     ..G..G......T.T..T..
Honeybee     .........AGTGA.A.T..     .........AGTGAAA.T..
Dobsonfly    .....G..............     .....G........T..G..
Froghopper   .....G..C......A....     .....G..C.....AA.A..
Kissing bug  ..G..C.....G........     ..G..C.....G..T..G..
Locust       ....................     ..............T..A..
Stonefly     .....G.....G........     .....G.....G..T..A..
Louse        .TT..A..............     .TT..A........A..G..
Collembola   .....A......TT......     .....A......TTA..G..
Daphnia      ..G..G..C...........     ..G..G..C.....C..A..
Artemia      ........C..G........     ........C..G..A..G..
Human        .....CA.C..G........     .....CA.C..G..A..A..
Turtle       ......A.C...AT......     ......A.C...ATA..A..
Chiton       .....G..............     .....G........A..A..
Snail        ....................     ..............A..A..
Earthworm    ..G..C.....G........     ..G..C.....G..A..A..
```

N4-N9153 is a 20 base primer with four-fold degeneracy at one site, and two-fold degeneracy at another site. It works well for many insects, and matches well with other Metazoa. It can be paired with N4-J8641 to amplify the middle portion of the *nad4* gene.

```
N4-J9172              10        20     without degeneracy:
(20mer)   5'-CGCTCAGGYTGRTACCCYCA  5'-CGCTCAGGYTGRTACCCYCA
Drosophila   ..T..C........A.....     ..T..C..T..A..A..T..
Mosquito     ....................     ........C..A.....T..
Silk moth    .....T........A.....     .....T..T..A..A..T..
Asp. beetle  ..T...........G.....     ..T.....C..A..G..C..
Flour beetle .....C........T.....     .....C..T..G..T..T..
Sawfly       ..A....A............     ..A....AT..A.....C..
Honeybee     ..A.TTTCACT...A.....     ..A.TTTCACTA..A..T..
Dobsonfly    ....................     ........T..A.....T..
Froghopper   ..T.TT..............     ..T.TT..T..G.....T..
Kissing bug  ..............G.....     ........C..A..G..C..
Locust       ..T...........A.....     ..T.....T..A..A..T..
Stonefly     ..T.................     ..T.....C..A.....T..
Louse        .....T........T..AA.     .....T..T..A..T..AA.
Collembola   .....TAA......T.....     .....TAAT..A..T..T..
Daphnia      ..T..G..............     ..T..G..T..G.....C..
Artemia      .....T........A.....     .....T..C..G..A..T..
Human        ..T..T.......TG.....     ..T..T..C..G.TG..T..
Turtle       ..T..TAT.....TA.....     ..T..TATT..G.TA..T..
Chiton       ..T..T..............     ..T..T..T..A.....T..
Snail        ..T..T........A.....     ..T..T..T..A..A..T..
Earthworm    ..T..T........G.....     ..T..T..C..A..G..C..
```

N4-J9172 is a 20 base primer with two-fold degeneracy at three sites. It works well with most insects, and should work with other Metazoa as well. Paired with either CB-N10608 or CB-N11010, it can be used to bridge the region between *nad4* and *cytb* in the ancestral arthropod genome.

```
N4L-N9629            10        20     without degeneracy:
(20mer)   5'-GTTTGTGAGGGWGYTTTRGG  5'-GTTTGTGAGGGWGYTTTRGG
Drosophila   ..A.....A..........     ..A.....A..A.C...A..
Mosquito     ........A..........     ........A..A.T...A..
Silk moth    ........A...T......     ........A..TTC...A..
Asp. beetle  ....................     ...........T.C...A..
Flour beetle ...........G........     ...........G.T...G..
Sawfly       ....C....T.........     ....C....T.T.T...A..
Honeybee     ..A......TT........     ..A......TTA.T...A..
Dobsonfly    ........A..G........     ........A..G.C...A..
Froghopper   ........A.....G.....     ........A..T.TG..G..
Kissing bug  ..............AA....     ...........T.CAA.G..
Locust       ....................     ...........T.C...G..
Stonefly     ........A..........     ........A..A.C...G..
Collembola   .CA.....A..G..A.....     .CA.....A..G.CA..A..
Daphnia      .C...C..............     .C...C.....A.C...A..
Human        .CC..C..A.C...GG....     .CC..C..A.CA.CGG.G..
Turtle       .CC.....A.C..GCA....     .CC.....A.CA.GCA.A..
```

N4L-N9629 is a 20 base primer with two-fold degeneracy at three sites. No reasonable alignment of this region was evident with Artemia or the non-arthropod invertebrates, so they are omitted from the alignment. It works very well with some insects, but is inconsistent. It can be paired with either N4-J8641 or N4-J8941 to bridge the *nad4* - *nad4l* junction. It completely overlaps primer N4L-J9648, below.

```
N4L-J9648            10        20       without degeneracy:
(21mer)   5'-ACCTAAAGCTCCCTCACAWAC  5'-ACCTAAAGCTCCCTCACAWAC
Drosophila   ............T........     ............T.....T..
Mosquito     .......A....T........     .......A....T.....A..
Silk moth    ........AA..T........     ........AA..T.....A..
Asp. beetle  .........A...........     .........A........A..
Flour beetle ...C...A.C...........     ...C...A.C........A..
Sawfly       C......A.A.A....G....     C......A.A.A....G.A..
Honeybee     .......A..AA.........     .......A..AA......T..
Dobsonfly    .........C..T........     .........C..T.....A..
Froghopper   T..C..CA.A..T........     T..C..CA.A..T.....A..
Kissing bug  ...C.TT..A...........     ...C.TT..A........A..
Locust       ...C.....A...........     ...C.....A........A..
Stonefly     T..C........T........     T..C........T.....A..
Collembola   C.....T..C..T......G.     C.....T..C..T.....TG.
Daphnia      C..............G...G.     C..............G..AG.
```

N4L-J9648 is a 21 base primer with two-fold degeneracy at one site. It works for a wide range of insects, but has little similarity to the homologous region in other Metazoa. It can be paired with CB-N10608 or CB-N11010 to amplify all of the *nad6* gene and the start of *cytb*. See also, N4-J9172, above. N4L-J9648 is essentially the complement of N4L-N9629.

```
CB-N10608              10        20          without degeneracy:
(23mer)       CCAAGTARTGAWCCAAARTTTCA  5'-CCAAGTARTGAWCCAAARTTTCA
Drosophila    ........................     .......A...T.....A.....
Mosquito      ..T.....................     ..T....A...T.....A.....
Silk moth     G.T.....G.....T.T......     G.T....GG..T..T.TA.....
Asp. beetle   ........A..............     .......AA..A.....G.....
Flour beetle  ..T.AA..A.....G........     ..T.AA.GA..T..G..G.....
Sawfly        ..T.A...A..............     ..T.A..AA..T.....G.....
Honeybee      .....A.T...............     .....A.T...T.....A.....
Dobsonfly     ..T.A......G...........     ..T.A..A...G.....A.....
Froghopper    ..T.A..................     ..T.A..G...T.....A.....
Kissing bug   .TT.A...A..............     .TT.A..AA..A.....A.....
Locust        ..T.............G......     ..T....A...T....GG.....
Stonefly      ..T.....G..............     ..T....AG..A.....G.....
Louse         ..T.A...ACT............     ..T.A..AACTT.....A.....
Collembola    ..C.A...G..............     ..C.A..GG..A.....A.....
Daphnia       ..T.AC.................     ..T.AC.A...A.....A.....
Artemia       ..C..G..A..C...........     ..C..G.GA..C.....G.....
Human         .....G.....G..G........     .....G.G...G..G..G.....
Turtle        ..T.....G..............     ..T....AG..T.....G.....
Chiton        .....A.....C...........     .....A.G...C.....A.....
Snail         ..G.....C.......T...C..     ..G....AC..A....TG..C..
Earthworm     ..C..A........GT.......     ..C..A.G...T..GT.G.....
```

CB-N10608 is a 23 base primer with two-fold degeneracy at three sites. The region it is based on is well conserved across Metazoa, but the primer is inconsistent in practice. When it works, it can be used to amplify the entire region between *nad4* and *cytb* by pairing it with either N4-J9172 or N4L-J9648 (see also, CB-N11010). CB-N10608 overlaps CB-J10621 by 16 nucleotides.

```
CB-J10621              10        20          without degeneracy:
(24mer)    5'-TTCAACAYRATGAAAYTTTGGWTC  5'-TTCAACAYRATGAAAYTTTGGWTC
Drosophila    .....G..................     .....G.TG......T.....A..
Mosquito      ...TG.T.................     ...TG.TTG......T.....A..
Silk moth     ...TT.T.........A.A.....     ...TT.TTG......TA.A..A..
Asp. beetle   .A......T...............     .A.....TT......C.....T..
Flour beetle  ......T.T.........C.....     ......TCT......C..C..A..
Sawfly        .AATTT..T...............     .AATTT.TT......C.....A..
Honeybee      .AATTATAT...............     .AATTATAT......T.....A..
Dobsonfly     ....G.T...............C..     ....G.TTG......T.....C..
Froghopper    A....G..................     A....G.TG......T.....A..
Kissing bug   ...TTT..................     ...TTT.TG......T.....T..
Locust        .....T..........C.......     .....T.TG......CC....A..
Stonefly      ...TG...................     ...TG..TG......C.....T..
Louse         ...TTATAT.............AG     ...TTATAT......T.....AAG
Collembola    ...C....................     ...C...TG......T.....T..
Daphnia       C....G...G..............     C....G.TGG.....T.....T..
Artemia       ...T.TT...............G..     ...T.TTTG......C.....G..
Human         C..CG.............C..C..     C..CG..TG......C..C..C..
Turtle        C..TG.C.................     C..TG.CTG......C.....A..
Chiton        A....TT...............G..     A....TTTG......T.....G..
Snail         C..T.TT..G..G...A.......     C..T.TTTGG..G..CA....T..
Earthworm     C..C.TT..........AC.....     C..C.TTTG......C.AC..A..
```

CB-J10621 is a 24 base primer with two-fold degeneracy at four sites. It sometimes fails, evidently because of mis-matches at the 5' end. CB-J10621 overlaps CB-N10608 by 16 nucleotides.

```
CB-J10933              10        20          without degeneracy:
(23mer)       GTTCTACCTTGAGGNCAAATRTC  5'-GTTCTACCTTGAGGNCAAATRTC
Drosophila    ...T...................     ...T.........A.....A..
Mosquito      ...T...................     ...T.........A.....A..
Silk moth     ..CT....C..............     ..CT....C.....T.....A..
Asp. beetle   ...T.G..C..............     ...T.G..C.....A.....G..
Flour beetle  ...T....C..............     ...T....C.....T.....A..
Sawfly        .....T.....G...........     .....T.....G..A.....A..
Honeybee      ..A.....A..............     ..A.....A.....A.....A..
Dobsonfly     .....T..C..............     .....T..C.....A.....A..
Froghopper    ..C.....A..............     ..C.....A.....T.....A..
Kissing bug   .....T..C..............     .....T..C.....A.....A..
Locust        ...T....A..............     ...T....A.....T.....A..
Stonefly      ..AT....C..............     ..AT....C.....A.....A..
Louse         .....T.....G...........     .....T.....G..G.....A..
Collembola    ..AT...................     ..AT.........A.....A..
Daphnia       .....T.....G.....G.....     .....T.....G..G..G..A..
Artemia       .....C..G..G...........     .....C..G..G..T.....G..
Human         ..C..C..G..............     ..C..C..G.....C.....A..
Turtle        ..C.....A..G...........     ..C.....A..G..C.....A..
Chiton        ..A.....C........G.....     ..A.....C.....A..G..A..
Snail         ..CT...................     ..CT.........A.....A..
Earthworm     .....G..C..G.........AG     .....G..C..G..A.....GAG
```

CB-J10933 is a 23 base primer with four-fold degeneracy at one site, and two-fold degeneracy at another site. This primer is a slight modification of the primer of the same name, listed in the Simon et al. (1994) review. It is quite reliable for most insects.

```
CB-N11010              10        20          without degeneracy:
(22mer)       TATCTACAGCRAATCCYCCYCA  5'-TATCTACAGCRAATCCYCCYCA
Drosophila    ......................     ..........A.....T..T..
Mosquito      .......T..............     .......T..A.....T..T..
Silk moth     ......................     ..........A.....C..T..
Asp. beetle   .TCT.........A........     .TCT......A..A..C..T..
Flour beetle  .G..G..T.....C........     .G..G..T..A..C..C..T..
Sawfly        ...TA.T..A............     ...TA.T..AA.....T..T..
Honeybee      ...TA.TT.A......A.....     ...TA.TT.AA.....A..T..
Dobsonfly     .G.....T..............     .G.....T..A.....C..T..
Froghopper    ....A.................     ....A.....A.....T..T..
Kissing bug   .........A............     .........AA.....T..T..
Locust        .......T..............     .......T..G.....T..T..
Stonefly      ......................     ..........A.....C..T..
Louse         CTCTA....A............     CTCTA....AG.....T..C..
Collembola    .TCT..................     .TCT......A.....T..T..
Daphnia       ............G........     ..........A..G..T..C..
Artemia       ....A........C........     ....A.....A..C..C..T..
Human         .G.....T.A.T.G........     .G.....T.AGT.G..T..T..
Turtle        .G.....T.A......A.....     .G.....T.AG.....A..T..
Chiton        .......T..............     .......T..A.....T..T..
Snail         G..T.....A...A.AG.....     G..T.....AA..A.AG..C..
Earthworm     .......T.....C..A.....     .......T..A..C..A..T..
```

CB-N11010 is a 22 base primer with two-fold degeneracy at three sites. It works for most insects, and can be used to bridge the gap between the *nad4l* and *cytb* genes when paired with either N4-J9172 or N4L-J9648.

```
CB-J11335            10        20     without degeneracy:
(20mer)       CATATTCAACCWGAATGRTA  5'-CATATTCAACCWGAATGRTA
Drosophila    ..C.................     ..C........A.....A..
Mosquito      ....................     ...........T.....A..
Silk moth     ..........C........     ...........C.....A..
Asp. beetle   ......A.............     ......A....A.....A..
Flour beetle  ..C.................     ..C........A.....A..
Sawfly        ......A.............     ......A....A.....A..
Honeybee      ......A.............     ......A....T.....A..
Dobsonfly     .....C..............     .....C.....A.....A..
Froghopper    ..C...A.............     ..C...A....A.....A..
Kissing bug   ....................     ...........A.....A..
Locust        ..C.................     ..C........A.....A..
Stonefly      .....C..............     .....C.....T.....A..
Louse         ..C.................     ..C........A.....A..
Collembola    ..C..C..............     ..C..C.....T.....A..
Daphnia       ....................     ...........T.....A..
Artemia       .....C.....G........     .....C.....G.....A..
Human         ..C..CA.G..C........     ..C..CA.G..C.....A..
Turtle        ......A.............     ......A....A.....A..
Chiton        .....C..G...........     .....C..G..A.....G..
Snail         ........G..C........     ........G..C.....A..
Earthworm     ......A.............     ......A....T.....A..
```

CB-J11335 is a 20 base primer with two-fold degeneracy at two sites. It works well with all insects tested so far, and can be paired with N1-N12067 or N1-N12242 to amplify this region in most insects that have the ancestral gene arrangement for *cytb* and *nad1*.

```
CB-N11526            10        20           without degeneracy:
(24mer)       TTCTACTGGTCGRGCTCCAATYCA  5'-TTCTACTGGTCGRGCTCCAATYCA
Drosophila    ...A....................     ...A........A........T..
Mosquito      ...A....................     ...A........G........T..
Silk moth     ......A...........G.....     ......A.....G.....G..T..
Asp. beetle   ......G........C........     ......G.....G..C.....T..
Flour beetle  .........A..T...........     .........A..T........T..
Sawfly        ....G.G.C..TTC..........     ....G.G.C..TTC.......T..
Honeybee      ...A.T.AA.T.TTTA..T.A...     ...A.T.AA.T.TTTA..T.AT..
Dobsonfly     ............C...........     ............C........T..
Froghopper    ...A........T...........     ...A........T........T..
Kissing bug   ...AG.A..A.....A........     ...AG.A..A..G..A.....T..
Locust        ............TTT.........     ............TTT......T..
Stonefly      ...G.....A..............     ...G.....A..A........T..
Louse         .........CAATCT....G....     .........CAATCT....G.T..
Collembola    .........G..............     .........G..G........T..
Daphnia       ......A..CTTT.....G.....     ......A..CTTT.....G..T..
Artemia       ......A.....T..C........     ......A.....T..C.....T..
Human         GCT.......T.TC....G.....     GCT.......T.TC....G..T..
Turtle        ...A..G...T.TC.C..G.....     ...A..G...T.TC.C..G..T..
Chiton        .........CT....A........     .........CT.A..A.....T..
Snail         ......C..A.AT..A..T.A...     ......C..A.AT..A..T.AC..
Earthworm     ......A......C.C........     ......A.....AC.C.....T..
```

CB-N11526 is a 24 base primer with two-fold degeneracy at two sites. It works for a wide range of insects, but is not as well conserved as some of the other *cytb* primers. Modifications are required for some insect groups.

```
N1-J11876              10        20          without degeneracy:
(23mer)   5'-CGAGGTAAAGTMCCWCGAACYCA  5'-CGAGGTAAAGTMCCWCGAACYCA
Drosophila   ........T.........G....     ........T..A..T...G.T..
Mosquito     ........................     ...........C..A.....T..
Silk moth    ........................     ...........A..A.....T..
Asp. beetle  .....A..T.........C....     .....A..T..C..A...C.C..
Flour beetle ..T..C.G...T...........     ..T..C.G...T..T.....C..
Sawfly       .....AT.T.A...C......A.     .....AT.T.AC..C.....TA.
Honeybee     .........A.........T...     .........A.A..T....TT..
Dobsonfly    ..T........T..G........     ..T........T..G.....T..
Froghopper   ..................T.....     ...........A..A..T..T..
Kissing bug  ..G..A.G..............     ..G..A.G...A..T.....C..
Locust       ........................     ...........A..A.....C..
Stonefly     .....A.GT..............     .....A.GT..A..T.....C..
Louse        ..T..A.G.A.T.......T.A.     ..T..A.G.A.T..T....TTA.
Collembola   ..G.....G.....C..G.....     ..G.....G..A..C..G..C..
Daphnia      ..G..A.G..............     ..G..A.G...C..T.....T..
Artemia      ..G..AT.G.AG..G..TCT...     ..G..AT.G.AG..G..TCTC..
Human        ..G..GT.T.CTGT.....T...     ..G..GT.T.CTGTT....TT..
Turtle       ..T..GT.T.AGG....G.....     ..T..GT.T.AGG.T..G..T..
Chiton       .....A..G..............     .....A..G..A..T.....T..
Snail        ..T..GT.TAC........TAA.     ..T..GT.TACA..A....TAA.
Earthworm    ..G..GT.T..TG..........     ..G..GT.T..TG.T.....C..
```

N1-J11876 is a 23 base primer with two-fold degeneracy at three sites. It works well for a wide range of insects.

```
N1-N12067              10        20          without degeneracy:
(23mer)   5'-AATCGWACTCCWTTTGATTTTGC  5'-AATCGWACTCCWTTTGATTTTGC
Drosophila   ........................     .....A.....T..........
Mosquito     ........................     .....A.....T..........
Silk moth    ........................     .....A.....T..........
Asp. beetle  ........................     .....T.....A..........
Flour beetle ........................     .....A.....T..........
Sawfly       ......G....C.........T.     .....AG....C.........T.
Honeybee     ...................AAT     .....A.....A........AAT
Dobsonfly    ........................     .....T.....T..........
Froghopper   ......T..................     .....TT....A..........
Kissing bug  ......T..................     .....TT....A..........
Locust       ..C..G..................     ..C..G.....T..........
Stonefly     ........................     .....T.....T..........
Louse        ......T.A........C..AT.     .....AT.A..T.....C..AT.
Collembola   .....GT..................     .....GT....A..........
Daphnia      ........C.....C........     .....T..C..A..C........
Artemia      ........A..C..C...C.A..     .....A..A..C..C...C.A..
Human        ..C.....C..C..C..CC....     ..C..A..C..C..C..CC....
Turtle       .....GG.A.....C..CC.AA.     .....GG.A..A..C..CC.AA.
Chiton       ..C...G.C........C.....     ..C..AG.C..T.....C.....
Snail        ......G.G..G..C..C.....     .....TG.G..G..C..C.....
Earthworm    ........C..G..C........     .....A..C..G..C........
```

N1-N12067 is a 23 base primer with two-fold degeneracy at two sites. It is quite reliable for most insects when paired with CB-J11335.

```
N1-N12242              10        20
(23mer)   5'-GTAGCCCAAACCATTTCTTATGA
Drosophila   ..G..T.....T...........
Mosquito     .....T.....T...........
Silk moth    .....T..G..T...........
Asp. beetle  ..T..T....TT...........
Flour beetle A.T..A.....G...........
Sawfly       ..T..T.....A...........
Honeybee     ...T.AAC..TA.....A.T...
Dobsonfly    .....T.....G...........
Froghopper   .....T.....T...........
Kissing bug  ..T..T..G..T...........
Locust       ..T..T.....T...........
Stonefly     ..T..T.....T...........
Louse        ...AGG...T.A...........
Collembola   .....T..G..T...........
Daphnia      ........G..TG.......C..
Artemia      .GG..A..G.....C..C.....
Human        ...........A..C..A.....
Turtle       ...........T..C..A..C..
Chiton       ..T..T..............C..
Snail        CAGCG......T..AAG......
Earthworm    ..T..T.....T.....A..C..
```

N1-N12242 is a non-degenerate 23 base primer that works for most insects. It overlaps N1-J12260 by 20 nucleotides.

```
N1-J12261              10        20           without degeneracy:
(24mer)   5'-AACTTCATAAGAAATAGTYTGRGC  5'-AACTTCATAAGAAATAGTYTGRGC
Drosophila   ........................     ...................T..A..
Mosquito     T........................     T..................T..A..
Silk moth    T........................     T..................C..A..
Asp. beetle  ................A.......     ................A.T..A..
Flour beetle C..............C.....T..     C..............C..T..T..
Sawfly       ...C...........T........     ...C...........T..T..A..
Honeybee     ..T....A.T.....TA..GTT.A     ..T....A.T.....TA.TGTT.A
Dobsonfly    C..............C........     C..............C..T..A..
Froghopper   T..C.....................     T..C..............T..A..
Kissing bug  T........................     T..................C..A..
Locust       T........................     T..................T..A..
Stonefly     ........................     ...................T..A..
Louse        ..T............T.A...CCT     ..T............T.AT..CCT
Collembola   ........................     ...................C..A..
Daphnia      T..C..G.......C.........     T..C..G.......C...C..G..
Artemia      T..C.....G..G..G.....T..     T..C.....G..G..G..C..T..
Human        G........T..G..T........     G........T..G..T..T..G..
Turtle       T.....G..T..G...........     T.....G..T..G.....T..G..
Chiton       T.....G........G........     T.....G........G..T..A..
Snail        G.T.......CTT.........CG     G.T.......CTT.....T..GCG
Earthworm    ...C..G..T..............     ...C..G..T........T..A..
```

N1-J12261 is a 24 base primer with two-fold degeneracy at two sites. It works well paired with LR-N13000 in species with the ancestral arthropod arrangement for the *nad1 - lsu* (16S) region. It has an almost complete overlap with N1-N12242.

```
N1-N12595              10        20              without degeneracy:
(26mer)   5'-GTWGCTTTTTTAACTTTATTRGARCG  5'-GTWGCTTTTTTAACTTTATTRGARCG
Drosophila   ..........................     ..A.................A..A..
Mosquito     ..........G.............     ..T........G........A..A..
Silk moth    ..........G.............     ..T........G........G..A..
Asp. beetle  ..............A..GC.....     ..T...........A..GC.A..A..
Flour beetle .....C.........C........     ..A..C.........C....A..A..
Sawfly       ...T.............A.......     ..TT..............A.A..G..
Honeybee     .....A......................     ..A..A..............A..A..
Dobsonfly    ..........G.............     ..A........G........G..A..
Froghopper   ....GA..C........GG.T.....     ..T.GA..C........GG.T..A..
Kissing bug  .....G...ACT......A.......     ..A..G...ACT......A.G..G..
Locust       ..................A.......     ..T..............A.G..G..
Stonefly     ..........................     ..A.................A..G..
Louse        .....A.....T..............     ..T..A.....T........A..A..
Collembola   ..G......A...T.C.TC.T.....     ..G......A...T.C.TC.T..G..
Daphnia      .....G.....C........T.....     ..A..G.....C........T..G..
Artemia      ..............AC..........     ..T...........AC....A..A..
Human        A.G..A..CC...TGC.TACC.....     A.G..A..CC...TGC.TACC..A..
Turtle       .....C.....C..CC..C.......     ..T..C.....C..CC..C.A..A..
Chiton       .....A.....C..............     ..A..A.....C........A..A..
Snail        .......A.A................     ..T....A.A..........A..A..
Earthworm    A....C..C.AC..AC..A.......     A.A..C..C.AC..AC..A.A..G..
```

N1-N12595 is a 26 base primer with two-fold degeneracy at three sites. It works well with many insects when paired with N1-J11876, amplifying the middle portion of the *nad1* gene. This primer is a slight modification of a primer of the same name in Simon et al. (1994).

```
LR-J12888              10        20           without degeneracy:
(23mer)   5'-CCGGTCTGAACTCARATCATGTA  5'-CCGGTCTGAACTCARATCATGTA
Drosophila   .....T.................     .....T........G........
Mosquito     .....T.................     .....T........G........
Silk moth    .....T.................     .....T........G........
Asp. beetle  .....T.A...............     .....T.A......G........
Flour beetle .....T.A...............     .....T.A......G........
Sawfly       .....T.................     .....T........G........
Honeybee     T..A.T.................     T..A.T........A........
Dobsonfly    .....T.................     .....T........A........
Froghopper   ........................     ..............G........
Kissing bug  ........................     ..............G........
Locust       ........................     ..............G........
Stonefly     ........................     ..............G........
Louse        T......T...............     T......T......A........
Collembola   ........................     ..............A........
Daphnia      ........................     ..............G........
Artemia      ...................C...     ..............G....C...
Human        ...................C...     ..............G....C...
Turtle       ...................C...     ..............G....C...
Chiton       ........................     ..............G........
Snail        .........-.............     .........-....G........
Earthworm    .-.............C.......     .-............GC.......
```

LR-J12888 is a 23 base primer with two-fold degeneracy at one site. It is a slight modification of LR-J12887 primer in the Simon et al. (1994) list, and is extremely versatile.

```
LR-N13000            10        20
(23mer)   5'-TTACCTTAGGGATAACAGCGTAA
Drosophila   ....T..................
Mosquito     ......................
Silk moth    ......................
Asp. beetle  ......................
Flour beetle ......................
Sawfly       .......G.............T.
Honeybee     ......................
Dobsonfly    ......................
Froghopper   ......................
Kissing bug  ......................
Locust       ......................
Stonefly     ......................
Louse        ......C..............T.
Collembola   ......A..............T.
Daphnia      .....C.................
Artemia      ......................
Human        .....C..............C..
Turtle       C....CC.............C..
Chiton       C....G.............AGTT
Snail        C...TC.............A...
Earthworm    C.................GC...
```

LR-N13000 is a non-degenerate 23 base primer. It works with most insects when paired with N1-J12260 to amplify and sequence the junction between these genes. It could be made more versatile if shortened by two or three residues at the 3' end.

```
LR-J13342            10        20          without degeneracy:
(24mer)   5'-CC-TTCGCACRGTCAAAATACYGC  5'-CC-TTCGCACRGTCAAAATACYGC
Drosophila   ..-..T.................     ..-..T....A.........T..
Mosquito     ..-....................     ..-.......G.........C..
Silk moth    ..-..T.T...............     ..-..T.T..A.........T..
Asp. beetle  ..-..A.................     ..-..A....G.........C..
Flour beetle ..-..T.................     ..-..T....A.........T..
Sawfly       ..-..T.................     ..-..T....A.........T..
Honeybee     ..-..T.T........C.......     ..-..T.T..A.....C....T..
Dobsonfly    ..-..T.................     ..-..T....A.........C..
Froghopper   ..-..T..........T......     ..-..T....A......T...T..
Kissing bug  ..-..A.......T.........     ..-..A....A..T.......T..
Locust       ..A..T.........T........     ..A..T....A....T.....C..
Stonefly     ..-..T.................     ..-..T....G.........C..
Louse        ..-..A.......-----.T..TG     ..-..A....A..-----.T.CTG
Collembola   ..-............T........     ..-.......A....T.....T..
Daphnia      ..-..A.........G........     ..-..A....A....G.....T..
Artemia      ..-......T.....GT.....A.     ..-......TG....GT....CA.
Human        ..-..T.......T.GGG......     ..-..T....G..T.GGG...C..
Turtle       ..-..T.......T.GG.......     ..-..T....G..T.GG....C..
Chiton       ..-..A.........GGG......     ..-..A....A....GGG...T..
Snail        ..-..T.........G.G......     ..-..T....A....G.G...T..
Earthworm    ..-..T.......T.GG.......     ..-..T....G..T.GG....C..
```

LR-J13342 is a 24 base primer with two-fold degeneracy at two sites.

```
LR-N13889            10        20          without degeneracy:
(23mer)   5'-ATTTATTGTACCTTKTGTATCAG  5'-ATTTATTGTACCTTKTGTATCAG
Drosophila   ........................     ..............G........
Mosquito     ........................     ..............G........
Silk moth    ..........T.............     ..........T...G........
Asp. beetle  ....T...................     ....T.........G........
Flour beetle ....T.A........A.......     ....T.A.......TA.......
Sawfly       T...GG..................     T...GG........G........
Honeybee     ....T...................     ....T.........T........
Dobsonfly    .A..T...................     .A..T.........T........
Froghopper   T........................     T.............T........
Kissing bug  .....A...................     .....A........T........
Locust       ....TC.T................     ....TC.T......T........
Stonefly     ........................     ..............G........
Louse        ....T.G....T.....C....C     ....T.G....T..T..C....C
Collembola   .GAATAGA............T..     .GAATAGA......T.....T..
Daphnia      T.....C.................     T.....C.......G........
Artemia      ...GT...................     ...GT.........T........
```

LR-N13889 is a 23 base primer with two-fold degeneracy at one site. Although this region is well conserved across arthropods, no reasonable alignment is evident for other Metazoa. It overlaps LR-J13900 by 12 residues.

```
LR-J13900            10        20          without degeneracy:
(23mer)   5'-TTTGATAAACYCTGATACAMAAG  5'-TGATAAACYCTGATACAMAAG
Drosophila   ...A......G............     .A......G........C...
Mosquito     ...A....................     .A......C........C...
Silk moth    ...A....................     .A......T........C...
Asp. beetle  ......T.................     ....T...C........C...
Flour beetle ...A..............T....     .A......C.......TA...
Sawfly       ...A....................     .A......C........C...
Honeybee     ...A..C.................     .A..C...C........A...
Dobsonfly    A..A....................     .A......C........A...
Froghopper   A..A..T.................     .A..T...C........A...
Kissing bug  A..A..T.................     .A..T...C........A...
Locust       ........................     ........C........A...
Stonefly     ........................     ........C........C...
Louse        ..CT.......G....G.....A     CT......CG....G..A..A
Collembola   ..CA.ATT.....A.........     CA.ATT..C..A.....A...
Daphnia      ....CAG.................     ..CAG...C........C...
Artemia      ...A....................     .A......C........A...
```

LR-J13900 is a 23 base primer with two-fold degeneracy at two sites. As noted for primer LR-N13889, which it overlaps by 12 residues, this region does not appear to be conserved outside of the Arthropoda.

```
SR-J14197            10        20          without degeneracy
(23mer)   5'-GTACAYCTACTATGTTACGACTT   5'-TACAYCTACTATGTTACGACTT
Drosophila   ........................     .....T.................
Mosquito     ........................     .....T.................
Silk moth    ....C......T...........     ....CT.....T...........
Asp. beetle  A..AG.T....T...........     A..AGTT....T...........
Flour beetle A..AG......T...C.......     A..AGC.....T...C.......
Sawfly       ..........T...........     .....T.....T...........
Honeybee     ....T.T................     ....TCT................
Dobsonfly    ..........T...........     .....C.....T...........
Froghopper   ...A.A.....T...........     ...A.A.....T...........
Kissing bug  ...AG......T...........     ...AGC.....T...........
Locust       ........................     .....T.................
Stonefly     ........................     .....C.................
Louse        ...A..T....G....T......     ...A.TT....G....T......
Collembola   .C.G.......T...........     .C.G.T.....T...........
Daphnia      ..........T...........     .....T.....T...........
Artemia      ...T.......T...........     ...T.C.....T...........
Human        ......T...C............     .....CT...C............
Turtle       ..GT..T...CT...........     ..GT.CT...CT...........
Chiton       C...T..................     C...TC.................
Snail        C...T..................     C...TT.................
Earthworm    T..........T...........     T....C.....T...........
```

SR-J14197 is a 23 base primer with two-fold degeneracy at one site. It is extremely versatile, and may be paired with SR-N14745 to amplify a major portion of the small subunit ribosomal RNA gene. It overlaps SR-N14182 by 16 residues.

```
SR-N14220            10        20     without degeneracy:
(20mer)   5'-ATATGYACAYATCGCCCGTC  5'-ATATGYACAYATCGCCCGTC
D. yakuba    ....................     .....T...C..........
Anopheles    .C..................     .C...C...C..........
Silkworm     ............T.......     .....T...T..T.......
Crioceris    ....................     .....T...C..........
Tribolium    ............T.......     .....C...T..T.......
Sawfly       ......C.............     .....TC..T..........
Honeybee     .AG......A..........     .AG..T...A..........
Dobsonfly    ....................     .....T...T..........
Froghopper   T...................     T....C...C..........
Kissing bug  T...................     T....C...T..........
Locust       G.G.................     G.G..T...C..........
Stonefly     .C..................     .C...C...C..........
Louse        T........A..........     T....T...A..........
Collembola   T...................     T....C...T..........
Daphnia      ..G.................     ..G..C...T..........
Artemia      ..G.A......-........     ..G.AC...T.-........
Human        GCGC.......C........     GCGC.T...C.C........
Turtle       G.GC.......C........     G.GC.C...C.C........
Chiton       .CG.................     .CG..T...T..........
Snail        C.G......A..........     C.G..C...A..........
Earthworm    .C..................     .C...C...C..........
```

SR-N14220 is a 20 base primer with two-fold degeneracy at two sites. This region is highly conserved and the primer can be used with LR-J12888 to amplify most of the large ribosomal subunit and the tRNA-val gene in most Metazoa, giving about a 1.4 kb fragment.

```
SR-J14610            10         20
(23mer)   5'-ATAATAGGGTATC-TAATCCTAGT
Drosophila   .............-..........
Mosquito     .............-..........
Silk moth    .............-..........
Asp. beetle  ....C........-..........
Flour beetle .............-..........
Sawfly       .............-..........
Honeybee     ...G.........-..........
Dobsonfly    .............-..........
Froghopper   .............-..........
Kissing bug  .....G....C..-......C...
Locust       .............-..........
Stonefly     ............AA..........
Louse        .............-..........
Collembola   ....C.....C..-.......G..
Daphnia      ....C........-.......G..
Artemia      C............-..........
Human        ...G.G.......-......C...
Turtle       ...G.G.......-......C...
Chiton       ...G.G.......-......C...
Earthworm    ...GAC.......-......GT..
```

SR-J14610 is a 23 base non-degenerate primer that works for most insects. It can be used to amplify and sequence across the control region (A+T rich region) in arthropods, when paired with an 'N' primer in tRNA-Ile or tRNA-Met. It may need to be modified near the 3' end for some arthropods or other Metazoa.

```
SR-N14745            10         20        without degeneracy:
(22mer)   5'-GTGCCAGCAGY-YGCGGTTANAC  5'-GTGCCAGCAGY-YGCGGTTANAC
Drosophila   ...........-..........     ..........T-C.......T..
Mosquito     ........TA.-..........     ........TAC-T.......T..
Silk moth    ...........-..........     ..........T-T.......T..
Asp. beetle  ...........-..........     ..........T-T.......G..
Flour beetle ...........-..........     ..........T-T.......G..
Sawfly       .........T.C..........     .........TTCC.......T..
Honeybee     TC.......A.-A..........    TC.......AT-A.......A..
Dobsonfly    .........T.-..........     .........TC-C.......T..
Froghopper   .........A.-..........     .........AT-C.......T..
Kissing bug  ...........-..........     ..........T-C.......T..
Locust       ...........-..........     ..........C-C.......T..
Stonefly     ........C..-..........     ........C.C-C.......T..
Louse        .........T.-A.........     .........TT-A.......A..
Collembola   ...........-..........     ..........T-C.......T..
Daphnia      ........C..-..........     ........C.T-C.......G..
Artemia      .........T.-......C....    .........TC-C.....C.T..
Human        ........CA.-......C....    ........CAC-C.....C.C..
Turtle       ........CA.-..........     ........CAC-C.......C..
Chiton       .........A.-..........     .........AC-T.......T..
Snail        ...........-...-..C....    ..........T-C..-..C.C..
Earthworm    ........T..-..........     ........T.C-C.......G..
```

SR-N14745 is a 22 base primer with four-fold degeneracy at one site, and two-fold degeneracy at two other sites. It is quite reliable for amplifying and sequencing much of the small ribosomal subunit, when paired with either LR-J13900 or SR-J14197.

## References


Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich D P, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465

Beard CB, Hamm DM, Collins FH 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* 2:103-124

Beckenbach AT, Robson SKA, Crozier RH. 2005. Single nucleotide +1 frameshifts in an apparently functional mitochondrial cytochrome b gene in ants of the genus *Polyrhachis*. *J. Mol. Evol.* 60:141-152

Boore JL, Brown WM 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata. Genetics* 138:423-443

Boore JL, Brown WM 1995. Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. *Genetics* 141:305-319

Boyce TM, Zwick ME, Aquadro CF. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* 123:825-836

Castro LR, Dowton M 2005. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). *Mol. Phylogenet. Evol.* 34:469-479

Clary DO, Wolstenholme DR 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* 22:252-271

Crease TJ 1999. The complete sequence and gene organization of the mitochondrial genome of the branchiopod crustacean, *Daphnia pulex* (Cladocera: Crustacea). *Gene* 233:89-99

Crozier RH, Crozier YC 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133:97-117

Dotson EM, Beard CB 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata. Insect Mol. Biol.* 10:205-215

Flook PK, Rowell CHF, Gellissen G 1995. The sequence, organization and evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* 41:928-941

Friedrich M, Muqim N 2003. Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castanaeum*. *Mol. Phylogenet. Evol.* 26:502-512

Lewis DL, Farr CL, Farquhar AL, Kaguni LS. 1994. Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Mol. Biol. Evol.* 11:523-538

Mindell, DP, Sorenson MD, Dimcheff DE. 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol. Biol. Evol.* 15:1568-1571

Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Ast JC, Yuri T 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Syst. Biol.* 48:138-152

Nardi F, Carapelli A, Fanciulli PP, Dallai R, Frati F 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations *Mol. Biol. Evol.*18:1293-1304

Perez ML, Valverde JR, Batuecas B, Amat F, Marco R, Garesse R 1994. Speciation in the *Artemia* genus: mitochondrial DNA analysis of bisexual and parthenogenetic brine shrimps. *J. Mol. Evol.* 38:156-168

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. *Ann. Entomol. Soc. Am.* 87:651-701

Shao R, Campbell NJH, Barker SC. 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Mol. Biol. Evol.* 18:858-865

Stewart JB, Beckenbach AT. 2003. Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*. *Mol. Phylogenet. Evol.* 26:513-526

Stewart JB, Beckenbach AT. 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopidae). *Genome* 48:46-54

Stewart JB, Beckenbach AT. 2006. Insect mitochondrial genomics. 2. The complete mitochondrial genome sequence of a giant stonefly, *Pteronarcys princeps*. Asymmetric directional mutation bias and conserved plecopteran A+T region elements. *Genome* (in press)

Terrett JA, Miles S, Thomas RH 1996. Complete DNA sequence of the mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Pulmonata). *J. Mol. Evol.* 42:160-168

Supplemental Material: Annu. Rev. Ecol. Evol. Syst. 2006. 37:545-79
doi: 10.1146/annurev.ecolsys.37.091305.110018
*Incorporating Molecular Evolution into Phylogenetic Analysis, and a New Compilation of Conserved Polymerase Chain Reaction Primers for Animal Mitochondrial DNA*
Simon, Buckley, Frati, Stewart, and Beckenbach

| **Primer** | **Sequence** |
|---|---|
| TI-N18 | TCCTATCAARRTAAYCCTTT |
| TI-J34 | GCCTGATAAAAAGGRTTAYYTTGATA |
| TM-N200 | ACCTTTATAARTGGGGTATGARCC |
| TM-J210 | AATTAAGCTACTAGGTTCATACCC |
| N2-J586 | CCATTTCAYTTYTGATTYCC |
| N2-N993 | GGTAAAAATCCTAAAAATGGNGG |
| TW-N1284 | ACARCTTTGAAGGYTAWTAGTTT |
| TW-J1301 | GTTAAWTAAACTAATARCCTTCAAA |
| TY-N1433 | GGCTGAATTTTAGGCGATAAATTGTAAA |
| C1-J1709 | AATTGGWGGWTTYGGAAAYTG |
| C1-N1738 | TTTATTCGTGGRAATGCYATRTC |
| C1-J2195 | TGATTCTTTGGWCACCCWGAAGT |
| C1-N2353 | GCTCGTGTATCAACGTCTATWCC |
| C1-J2756 | ACATTCTTTCCTCARCAYTT |
| C1-N2776 | GGTAATCAGAGTATCGWCGNGG |
| TL-J3033 | TCTAATATGGCAGATTAGTGCA |
| TL-J3043 | GGCAGATTAGTGYAATGRATTTAA |
| C2-N3389 | TACTCATARGATCARTATCAYTG |
| C2-J3399 | ACAATTGGTCAYCAATGATAYTG |
| C2-J3624 | ACTCCTGGACGATTAAAYCA |
| C2-N3665 | CCACAAATTTCTGAACATTG |
| TK-J3790 | CATTAGATGACTGAAAGCAAGTA |
| TK-N3796 | ACTATTAGATGGTTTAAGAG |
| A8-N4061 | GAGAATAAGTTWGTTATCATTTTCA |
| A6-J4463 | TTTGCCCATCTWGTWCCNCAAGG |
| A6-N4552 | ATGTCCWGCAATYATATTWGC |
| C3-J4792 | GTTGATTATAGACCWTGRCC |
| C3-N4908 | CGAGTTAYATCTCGTCATCATTG |
| C3-N5460 | TCAACAAAATGTCARTAYCA |
| C3-J5470 | GCAGCTGCYTGATAYTGRCA |
| N3-N5731 | TTAGGGTCAAATCCRCAYTC |
| N3-J5747 | CCATTTGAATGTGGRTTTGAYCC |
| TN-J6155 | TTTAATTGAARCCAAAAAGAGG |
| TN-N6160 | TCAATTTTRTCATTAACAGTGA |
| TN-J6172 | AGAGGTATATCACTGTTAATGA |
| TF-N6384 | TATATTTAGAGYATRAYAYTGAAG |
| TF-J6400 | TAACATCTTCAATRTYATRCTCT |
| N5-J7077 | TTAAATCCTTWGARTAAAAYCC |
| N5-N7211 | TTAAGGCTTTAYTATTTATRTGYGC |
| N5-J7572 | AAAGGGAATTTGAGCTCTTTTWGT |
| N5-N7793 | TTAGGTTGRGATGGNYTAGG |
| N5-J7806 | GAMACAARACCTAACCCATCYCA |
| N4-N8487 | TATCAGSTAATATRGCWGCTCC |
| N4-J8641 | CCAGAAGAACATAANCCRTG |

| Primer | Sequence |
|---|---|
| N4-N8727 | AAATCTTTRATTGCTTATTCWTC |
| N4-J8941 | GAAACAGGAGCCTCAACATGWGC |
| N4-N9153 | TGAGGTTATCAACCNGARCG |
| N4-J9172 | CGCTCAGGYTGRTACCCYCA |
| N4L-N9629 | GTTTGTGAGGGWGYTTTRGG |
| N4L-J9648 | ACCTAAAGCTCCCTCACAWAC |
| CB-N10608 | CCAAGTARTGAWCCAAARTTTCA |
| CB-J10621 | TTCAACAYRATGAAAYTTTGGWTC |
| CB-J10933 | GTTCTACCTTGAGGNCAAATRTC |
| CB-N11010 | TATCTACAGCRAATCCYCCYCA |
| CB-J11335 | CATATTCAACCWGAATGRTA |
| CB-N11526 | TTCTACTGGTCGRGCTCCAATYCA |
| N1-J11876 | CGAGGTAAAGTMCCWCGAACYCA |
| N1-N12067 | AATCGWACTCCWTTTGATTTTGC |
| N1-N12242 | GTAGCCCAAACCATTTCTTATGA |
| N1-J12261 | AACTTCATAAGAAATAGTYTGRGC |
| N1-N12595 | GTWGCTTTTTTAACTTTATTRGARCG |
| LR-J12888 | CCGGTTTGAACTCARATCATGTAA |
| LR-N13000 | TTACCTTAGGGATAACAGCGTAA |
| SR-J13342 | CCTTCGCACRGTCAAAATACYGC |
| LR-N13889 | ATTTATTGTACCTTKTGTATCAG |
| LR-J13900 | TTTGATAAACYCTGATACAMAAG |
| SR-J14197 | GTACAYCTACTATGTTACGACTT |
| SR-N14220 | ATATGYACAYATCGCCCGTC |
| SR-J14610 | ATAATAGGGTATCTAATCCTAGT |
| SR-N14745 | GTGCCAGCAGYYGCGGTTANAC |