Sequencing and comparing whole mitochondrial genomes of animals

Jeffrey L. Boore[1,2], J. Robert Macey[1], and Mónica Medina[1]

[1] Evolutionary Genomics Department

DOE Joint Genome Institute and Lawrence Berkeley National Lab

2800 Mitchell Drive

Walnut Creek, CA 94598

[2] Department of Integrative Biology

University of California Berkeley

3060 Valley Life Sciences Building

Berkeley, CA 94720

Corresponding author: Jeffrey Boore

Phone: 925-296-5691

Fax: 925-296-5620

E-mail jlboore@lbl.gov

Running title: Mitochondrial Genomics

I. ABSTRACT

Comparing complete animal mitochondrial genome sequences is becoming increasingly common for phylogenetic reconstruction and as a model for genome evolution. Not only are they much more informative than shorter sequences of individual genes for inferring evolutionary relatedness, but these data also provide sets of genome-level characters, such as the relative arrangements of genes, that can be especially powerful. We describe here the protocols commonly used for physically isolating mtDNA, for amplifying these by PCR or RCA, for cloning, sequencing, assembly, validation, and gene annotation, and for comparing both sequences and gene arrangements. On several topics, we offer general observations based on our experiences to date with determining and comparing complete mtDNA sequences.

I. BACKGROUND

Mitochondria are subcellular organelles of nearly all eukaryotes, descended from alpha-proteobacteria that took up residence inside an early member of the eukaryotic lineage (Lang, Gray, and Burger, 1999). They still contain their own, much diminished genomes and have systems for transcription, message processing, and translation that are separate from those of the cytoplasm. For animals, these mitochondrial genomes are almost always circular (for exceptions in the Cnidaria see Bridge et al., 1992; in a crustacean see Raimond et al., 1992) and usually contain the same set of 37 genes, encoding 13 proteins, two rRNAs, and 22 tRNAs (see Boore, 1999; for some exceptions see Wolstenholme et al., 1987; Okimoto et al., 1991; Hoffmann, Boore, and Brown, 1992; Beagley et al., 1995; Keddie, Higazi, and Unnasch, 1998; Beagley, Okimoto, and Wolstenholme, 1998; Beaton, Roger, and Cavalier-Smith, 1998; Yokobori et al., 1999, 2003; Armstrong, Blok, and Phillips, 2000; Le et al., 2000; Nickisch-Rosenegk, Brown,

and Boore, 2001; Helfenbein et al., 2004). Typically, they are about 16 kb in size, and so are very gene-dense, and without introns except in cnidarians (Beagley et al., 1995; Beagley, Okimoto, and Wolstenholme, 1998; Beaton, Roger, and Cavalier-Smith, 1998), which also contain one extra gene, a homolog to bacterial *mutS*. Some mtDNAs have all genes on one strand; for others they are distributed between both. In the few cases where it has been studied, transcription produces a single large transcript for each DNA strand which is then enzymatically cut into (mostly) gene specific RNAs (see Clayton, 1992).

There are several merits to comparing these diminutive genomes. Their small size and compact arrangements facilitate broad comparisons for many animals. Comparisons can include the homologous genes found in the mtDNAs of plants, protists, and fungi, and in the genomes of prokaryotes (Lang et al., 1997; Gray et al., 1998; Gray, 1999; Adams and Palmer, 2003). Their (usually) circular structure enables physical isolation from nuclear DNA. Their biochemistry is relatively well understood and they are known to play important roles in cellular metabolism (e.g., Nieminen, 2003), development (e.g., Yost et al., 1995; Krakauer and Mira, 1999), aging (Nagley and Wei, 1998), and human disease (Wallace, 1999). Their products interact with those from hundreds of nuclear genes, inviting studies of co-evolution with these interacting factors (Wu et al., 2000). Many of the processes of genome evolution can be studied in these relatively simple systems, including genome rearrangements (Boore, 2000), tRNA editing (Lavrov, Brown, and Boore, 2000), tRNA gene "identity theft" (Rawlings, Collins, and Bieler, 2003), and the causes of mutational biases and their effects on amino acid substitution patterns (e.g., Helfenbein, Brown, and Boore, 2001).

Mitochondrial genome comparisons have successfully addressed a broad range of phylogenetic questions. Rapidly evolving portions of non-coding DNA are used for forensic

identifications (Budowle et al., 2003) and addressing population structure (e.g., Nyakaana, Arctander, and Siegismund, 2002). Although mtDNA sequences generally evolve more rapidly than those of nuclear genes (Brown, George, and Wilson, 1979; Gissi et al., 2000), leading some to question their resolving power, they have produced robust phylogenies even at very deep levels (e.g., Helfenbein et al., 2004). Whole mtDNA sequence comparisons are much more powerful for phylogenetic reconstruction than single gene comparisons (Ingman et al., 2001; Parsons and Coble, 2001; Boore, Medina, and Rosenberg, 2004; Macey et al., 2004). Complete mitochondrial genome sequences also provide a set of "genome-level characters" (Nikaido, Rooney, and Okada, 1999; Schmitz, Ohme, and Zischler, 2001), such as RNA secondary structures (Macey et al., 1997c; Macey, Schulte, and Larson, 2000), modes of control of replication and transcription (Clayton, 1992), mtDNA physical structures (Bridge et al., 1992), and especially the relative arrangements of genes, that can be a very reliable indicator of common ancestry (see below).

Figure 1 and Table I about here

For these reasons, and because high-throughput genome sequencing is becoming a mature technology, comparing complete mtDNA sequences is becoming increasingly common. In no other part of the genome could one so easily obtain the sequences of 37 unambiguously orthologous genes so densely packed. The rate of production of complete mtDNA sequences is increasing exponentially and, as of June, 2004, GenBank holds 464 complete animal mtDNA sequences. Although taxonomic sampling is highly biased at present (Figure 1), we are moving toward a thorough sampling across life of this small genome for better understanding of the evolution of genomes and organisms. (Table I lists some of the resources on the web for

learning more about mitochondrial systems.) It is imperative that we streamline the processes of acquiring, analyzing, and comparing these data.

II. MAKING THE TEMPLATES

Differences in body structures among various animal groups makes it impossible to comprehensively generalize the best tissues for isolating mtDNA. Eggs, gonads, muscle, liver, and brain are commonly used, generally in this order of preference, although other tissues are often acceptable. All methods are most reliable when using fresh tissue, although tissues frozen (quickly, if possible) at –80 degrees are commonly used. Working with tissues stored in ethanol is variably successful, but those in formalin have never yielded useful DNA in our hands [but see Kearney and Stuart (2004) for a case with obtaining small fragments from formalin fixed samples.].

There are numerous methods, including the use of several commercial products, for isolating total DNA from tissue that are described elsewhere. For study of any particular region of the genome, the next step must be to physically isolate that portion of the DNA. For the case of mtDNA, there are several methods in common usage: (1) Physically isolating mtDNA from nuclear DNA by exploiting the differential masses of the nucleus and organelle and/or the differing properties of linear vs. circular DNA; (2) Creating and identifying clones that contain copies of the mtDNA; or (3) Generating identical copies of the mtDNA in high molarity by long-PCR or rolling circle amplification.

1. Physical isolation

It is possible to generate about (nominally, and with high variance) 1 μg of mtDNA per gram of tissue using standard cell fractionation techniques (see Protocol 1). Especially if this DNA is ever to be used for PCR, it is critical to minimize the potential for contamination between samples or otherwise. Briefly, the tissue is disrupted using either a dounce homogenizer or a tissuemizer (Tekmar, Inc.), trying to maximize the breaking of fibrous matrix and cell membranes while minimizing the breaking of nuclear and mitochondrial membranes. This step may need optimizing, since too vigorous treatment will break more nuclei, reducing the amount of nuclear DNA removed by centrifugation, and too gentle treatment will incompletely lyse cells, which will pellet in this centrifugation, so removing mitochondria and reducing yield. It is critical to minimize enzymatic damage to DNA by working quickly, maintaining dilute solutions, and keeping reagents and materials cold. Nuclei are pelleted by low speed centrifugation, but nuclear DNA from any that have been inadvertently lysed will contaminate the supernatant (an amount typically larger than the isolated mtDNA itself after ultracentrifugation). It is optional at this point to layer mitochondria at the interface of a 1.5-to-1 M sucrose step gradient; this is most useful if a large amount of tissue was used. Mitochondria are pelleted and lysed, then the mtDNA is separated from any contaminating nuclear DNA by centrifugation in the presence of CsCl (to produce the primary gradient) and propidium iodide (PI) in an ultracentrifuge. The PI intercalates between base pairs causing unwinding of the DNA helix, which is physically resisted by circular, but not linear DNA, since the latter can unwind without limit. This renders circular DNA more dense than linear, which seems counterintuitive, but the dye binding lessens the mass since it introduces more "space" than weight. Detailed explanation and a description of the critical parameters for the ultracentrifugation can be found under sections for separating plasmids from genomic DNA in Sambrook and Russell (2001) and Ausubel et al. (2004). The mtDNA

fraction will be a few mm below the nuclear band. After collecting this with a needle, the PI is removed by extraction with butanol, then the CsCl removed by dialysis or ultrafiltration.

Figure 2 about here

This process requires relatively large amounts of tissue, a great effort, and often extensive optimization. However, it can be very reliable, especially when it can be optimized for a single tissue type. Figure 2 (lanes 2, 3, and 7) shows the results of a physical purification from the very mtDNA-rich eggs of *Xenopus*.

2. Cloning

The optimal insert size for plasmids is too small, and for fosmids is too large to accommodate most mtDNAs. The vector of choice is a replacement-type phage vector optimal for a mtDNA-sized insert such as EMBL3 or EMBL4 (available from Stratagene and others). The greatest problem with cloning an entire mtDNA into a phage is that one must begin with a single break, ordinarily generated by cutting with a restriction enzyme recognizing the mtDNA at only one site and compatible with (or modifiable to be compatible with) a site in the polylinker of the vector. This can be determined by constructing a restriction enzyme map of isolated mtDNA, if it is very pure, by radiolabeling fragments (Protocol 2), or by doing a Southern blot (Southern, 1975) of total DNA using (normally heterologous) mtDNA probes. Detailed protocols for this and other molecular techniques not individually referenced here can be found in Sambrook and Russell (2001) and Ausubel et al. (2004). Briefly, cloning into phage requires ligation of the vector and insert, packaging into infectious particles, transfecting bacteria, and plating, where the clones form plaques on a lawn of growing bacteria. If the library is made from isolated mtDNA, then it is normally sufficient to randomly pick a few clones and test each

for a mtDNA insert. If the library is made from total DNA instead, then this would be inefficient, and one would ordinarily probe plaque lifts to identify the correct clone(s). The best probe DNA for this would be short PCR fragments amplified from the mtDNA itself, so that the hybridization can use very stringent conditions, although a heterologous probe from other mtDNAs can work well. In either case, the candidate clone should be verified as containing the entire mtDNA by checking that the restriction enzyme map made from DNA produced from the clone, ordinarily visualized on an agarose gel stained with ethidium bromide (i.e., rather than by radiolabeling), matches that determined for the native mtDNA.

3A. PCR amplification

Because of the difficulties in purifying and/or cloning mtDNAs, many have adopted PCR amplification from whole genomic DNA extractions. Until a few years ago, PCR could amplify only fragments under a few kb, but now techniques are available for amplifying much longer regions (Cheng et al., 1994). Although we have been able to amplify an entire mtDNA from a single set of primers (Boore, Medina, and Rosenberg, 2004), it is a more efficient strategy with more common success to amplify each in two or three overlapping portions (Figure 3). In some cases we use primers designed to match conserved regions for long-PCR. However, it is generally more effective to amplify and determine a short sequence, then make primers of perfect match to the mtDNA and facing "out" from the fragment, since the success rate for long-PCR is lower than for shorter fragments, at least in our hands, and this eliminates the variable of the extent of primer matching. In either case, where mtDNAs have large-scale gene rearrangements, it is often necessary to try many combinations since it is impossible to tell *a priori* which primers are opposed on the mtDNA and of acceptable separation. In cases of

difficulty, we sometimes amplify and sequence one portion, nominally half, then make specific primers to amplify the remainder.

Figure 3 about here

We routinely use either Takara LA (Takara Bio, Inc.) or rTth-XL polymerase (Perkin-Elmer) essentially according to suppliers' instructions. [Other enzymes are available, including Elongase (Invitrogen) and Herculase (Stratagene), with which we have less experience.] Optimization of reaction conditions is often required, especially the magnesium concentration and primer annealing temperature. It is important that sufficient time be allowed for the extension step, especially in later cycles, so we typically start with one minute per kb to amplify for the first 15 cycles, then use the "autoextend" feature of the PCR machine to lengthen this step by an additional 15 seconds per cycle, usually to a total of 37 cycles.

Although there are more complex methods available, we find that it is sufficient to estimate primer annealing temperature by this simple formula: (2 x the number of A's or T's) plus (4 x the number of G's or C's) minus 5. The two primers should be as close as possible to the same estimated annealing temperature. Primer sequences should avoid long runs of homopolymers, be 40-60% G+C, and not have lengthy sequences at their 3' ends that are in reverse complement to one another. When designing primers to conserved regions, start by aligning amino acid sequences of other animals, being sure to include those distant enough to see saturation of substitutions for non-essential portions of the genes. Choose portions about eight amino acids long that are well conserved and about the desired distance apart. Favor regions that include amino acids that are of minimum possible codon variation, such as tryptophan and methionine, and avoid those that can be coded by many variations, such as serine and leucine. Primers can be made that are degenerate for all possible codon possibilities or that simply use the codons

expected to be most common for the amino acids; each strategy seems about equally likely to be successful. It is most important that there are no mismatches at the 3' end, so we recommend ending a primer on a second codon position of a universally conserved amino acid. Although these are useful guidelines, in practice, some primers work well that do not fully conform, and others fail even when conforming to all of these rules, for reasons that are not obvious. Table II lists a set of primers with which we have been routinely successful at amplifying short fragments from many animal mtDNAs spanning the diversity of the Metazoa.

Table II about here

One of the common problems we've experienced is amplifying regions that include the large non-coding region, sometimes called the "control region", which generally contains signals for origin and termination of replication. We speculate that in some cases, the polymerase used for the PCR may respond to the replication termination signal for the mtDNA, but there is no obvious similarity among these difficult regions either for sequence, base composition, or potential secondary structures. We have sometimes overcome difficulties by switching to a different polymerase or by reducing the size of the region to amplify by determining flanking sequences. Another effective strategy can be to use the greatly diluted product of an unsuccessful (as judged by having either multiple bands or nothing at the level of detection on the gel) PCR as template for a subsequent amplification with primers that are internally nested. Not uncommonly, using less of any template DNA results in a higher success rate of amplification, since this also dilutes any impurities that might be inhibitory; while amount of template DNA added can, at best, boost amplification product proportionally, inhibitory elements affect the reaction product exponentially.

3B. Rolling circle amplification

RCA is a recently developed technique for producing *in vitro* long, double-stranded, multiple tandem copies of circular DNA molecules (Lizardi et al., 1998; Fire and Xu, 1995; Liu et al., 1996). The technique is dependent on use of a DNA polymerase with high processivity and strong strand displacement activity, such as that from Phi29 bacteriophage (Lizardi et al., 1998). This enzyme allows extension of >70,000 bases from a single priming event (Blanco et al., 1989) and has high fidelity, yielding only 1 error in 10^7-10^8 bases (Estaban et al., 1993). After heat denaturation, multiple primers (typically random hexamers) are annealed, and then extended simultaneously at constant temperature by the polymerase. For a circular DNA molecule—like a mitochondrial genome—as each growing strand reaches its origin, it displaces itself, spooling off long, single stranded, tandem copies. These are converted during the RCA reaction into double stranded form by further primer annealing and extension of the complementary strand. RCA has been used to augment the signal from oligonucleotide probes (Baner et al., 1998), to amplify plasmid templates for high-throughput DNA sequencing (Dean et al., 2001; Detter et al., 2002), and even to amplify large genomes (Detter et al. 2002; Dean et al., 2002).

It is important to note that linear DNA undergoes strand displacement amplification in this same reaction, and that this is an isothermal reaction at low temperature, so there is limited specificity of amplification. Nonetheless, we have had success using six of the oligonucleotides to *rrnS*, *rrnL*, and *cox1* (Table II) on crude DNA preparations. In order to prevent exonucleolytic activity on the oligonucleotides, the two most 3' bonds must be phosphorothiol bonds rather than phosphodiesters. We have also successfully amplified mtDNAs that have been purified using ultracentrifugation as above by RCA using random hexamers. Although these

mtDNAs were presumably pure, they were of such small quantity that nothing was visible below the nuclear DNA band with UV irradiation of the CsCl gradient. Figure 2 (lanes 5 and 6) show one result using purified *Xenopus* mtDNA. For RCA using random hexamer primers, we use the protocols available with the kit (Amersham, Molecular Staging, or Epicentre); Protocol 3 describes the method we use for RCA with specific primers. The best measure of RCA success at amplifying an organelle genome is the ratio of DNA appearing in bands vs. smearing on a gel following digestion; this correlates well with the proportion of organelle DNA found in random sequencing reads. We have successfully produced several complete mtDNA sequences from random shotgun plasmid libraries (below) produced from these RCA products.

While both approaches have been successful, each has shortcomings and advantages. With specific primers, we can use total DNA preparations, but may be limited by endogenous enzyme inhibitors or by the ability to accurately design conserved primers. Further, several commercially available kits do not provide pure Phi29 enzyme, but a reaction mix that contains both enzyme and random hexamers. (Another problem with these kits is that there is nearly always amplification in an attempted negative control reaction, presumably due to contamination of the enzyme mix with DNA from the Phi29 plasmid clone.) Alternatively, although randomly-primed RCA requires highly purified mtDNA, it must be available only in miniscule amounts. Protocols to date for purifying mtDNA have emphasized quantitative recovery as has been necessary for direct cloning or restriction enzyme analyses, but RCA enables the use of alternatives that would generate only minute amounts of pure mtDNA.

Work in progress is attempting to increase the specificity of amplification even at the low RCA temperatures by modification of reaction conditions. Other experiments are testing the potential for transposon-mediated cloning of mtDNAs from total DNA preparations, purification

of mtDNAs using biotinylated primers matching mtDNA sequences coupled to streptavidin-coated magnetic beads, and isolating intact mitochondria using a Florescence Activated Cell Sorter (FACS).

III. SEQUENCING THE TEMPLATES

To date, no techniques are available for reading more than about 1 kb of DNA sequence in a single reaction, so whether the template is purified mtDNA, a mtDNA-containing phage clone, a long-PCR product, or the result of an RCA reaction, the task must be broken into smaller components by one of two methods: (1) The DNA to be sequenced is physically broken into smaller pieces which are cloned for "shotgun" sequencing (below), then the sequence that is determined for each is assembled based on overlap or (2) primers are designed to incrementally walk through a longer fragment. These methods and their variations will be described separately.

DNA is broken into smaller fragments by one of three methods. Restriction enzymes can cut the DNA into precise and reproducible fragments, but these vary greatly in size and can cut only at specific points, which may not be conveniently distributed. To minimize these problems, it is common to use a four base-recognizing enzyme, expected to cut about every 256 nucleotides, but under conditions that result in only partial digestion. This is difficult to control and requires optimization of conditions. Secondly, the DNA can be sonicated, that is, broken by intense sound waves using a specially constructed device. Breaks are random, but the distribution of fragment sizes is wide, so a very large amount of starting material is required. The best method is generally to break the DNA by driving it repeatedly through a narrow aperture, like the one presented in a HydroShear device (Gene Machines; protocol at <http://www.jgi.doe.gov/prod/SCLIB.html>). The DNA stretches as it passes through under

high pressure, and breaks if it is longer than a size specified by the pressure, typically to a size between 1.5 and 3 kb for mtDNAs. Figure 4 shows the mechanism of action and a typical result. [A common misconception is that we intend to sequence completely through these clones, leading to the comment that they are too large; actually, the sequencing reads from each end of a clone do not overlap, but rather we rely on reads from many clones for the contig assembly (see below).]

Figure 4 about here

In our process at JGI, after enzymatic end repair and electrophoretic size selection, these fragments are ligated into pUC18 and transformed into *E. coli* DH10b to create plasmid libraries. These are plated and grown overnight, then an automated colony picker (Genetix) is used to select colonies into 384-well plates of LB with 10% glycerol. These are incubated overnight in a static incubator, without shaking or enhanced aeration, then a small aliquot is processed robotically through plasmid amplification using RCA, sequencing reactions, reaction clean up using SPRI (Elkin et al., 2002) and processing on an automated capillary DNA sequencer. Detailed protocols are available at <http://www.jgi.doe.gov/Internal/prots_index.html>. Since these sequenced fragments are from random starting points, we determine at least 10 times the amount of sequence in the target template, then assemble these sequencing reads together for a complete mtDNA sequence (see below).

Another option is to primer walk through long PCR fragments or clones. Although there are many alternatives possible, we load PCR reactions onto ultrafiltration spin columns (Ultrafree 30,000 NMWL, Millipore), fill with water to 400 μl, and spin according to supplier's instructions through three repetitions. This removes primers, buffer, and unincorporated nucleotides and provides template directly for sequencing reactions. After each sequencing

reaction, a primer is selected for the next round until a high quality, complete sequence is determined from each strand. Primers are typically 18 nucleotides in length, although can be made longer to increase annealing temperature if necessary. In some cases, when sequencing multiple, related mtDNAs, researchers have made a series of primers to conserved regions to streamline the process (Sorenson et al., 1999; Inoue et al., 2001)

IV. ASSEMBLY OF THE SEQUENCING READS

The raw sequencing reads are first processed with PHRED (Green, 1996), which generates chromatograms, base calls, and quality scores. The latter are expressed as "Q values" that correspond to the probability of error according to the scale that Q10 = 1/10, Q20 = 1/100, Q30 = 1/1000, etc. Each quality file is linked to the corresponding base call file and these can be moved between platforms. Although there is other commercial software for generating chromatograms and base calling, PHRED's quality scores are an outstanding feature.

If the method of choice is primer walking, then assembly becomes trivial and can be done manually, since each subsequent sequencing read begins immediately beyond the primer annealing site in the previous read. The more complex issues are for the assembly of a large collection of sequencing reads from random clones. The most commonly used computer programs for this are Sequencher (Gene Codes) or PHRAP (Green, 1996). In most cases, the first step is to trim, to remove low quality sequence at the end of each read and to remove the small amount of vector sequences (part of the multiple cloning site) at the beginning of each read. This can be done automatically in either system. Details for the use of these particular programs can be found at <http://www.genecodes.com/> and <http://www.phrap.org/>.

Common difficulties are gaps remaining either from having an insufficient number of sequencing reads, from cloning biases, or from misassemblies due to repeated sequences. Gaps can be closed by directed approaches, either by PCR amplifying and sequencing the missing portions or, if a plasmid clone can be seen to span the gap, by primer walking with this clone as a template. Misassemblies due to repeats often appear as unusually deep coverage of a region by sequencing reads and a pattern where reads from opposing ends of clones are closer together than the typical clone size. This can be definitively resolved by using Southern hybridizations (Southern, 1975; Sambrook and Russell, 2001; Ausubel, 2004) or PCR amplifications from unique, flanking sequences.

All assemblies and sequence quality should be verified by eye using either Sequencher or CONSED (Green, 1996). The former is available for the Macintosh and is somewhat simpler to use; CONSED (as well as PHRED and PHRAP) is based in Unix, but has more features, especially for resolving misassemblies (and all three are free of charge). Even if the assembly was generated in PHRAP, the entire set of files, including quality scores, can be imported into Sequencher for verification and viewing of this assembly. Throughout the assembly there should be significant overlap of reads and multiple reads of high quality. One frequent problem is in gap handling for the consensus sequence, especially if multiple sequences of low quality are included.

If any portion of the sequence was generated from PCR fragments, it is critical to remove the sequences of the primers from the ends before assembly, since these may not exactly match the mtDNA sequence. If the mtDNA is circular, then any assembly will show it arbitrarily linearized, with some sequence repeated on each end of the assembly. It is critical to identify this and to trim one end back to unique sequence; the use of a dot matrix sequence identity plot, such as is available in MacVector (Accelrys) or other packages can help.

An iterative approach at verifying sequence quality and assembly is sometimes needed, where one pays particular attention to deficiencies discovered during gene annotation (below), such as a frame shift or stop codon within a coding region or mismatches in paired nucleotides in tRNA genes, to ensure that no error was generated. To the best of our knowledge, there is no software available that is effective in allowing visualization of many complex features while viewing assembled sequences, so this is a highly manual and iterative process.

V. ANNOTATION OF THE GENES

Identifying tRNA genes is the most challenging aspect of annotating animal mtDNAs, since there is little sequence similarity except among closely related animals, many features common to cytoplasmic tRNAs are absent, and many are of aberrant secondary structure. Figure 5 shows a labeled schematic diagram of a tRNA. Universally present is a seven member anticodon loop, a five base pair anticodon stem, and a seven member acceptor stem. Either stem may contain some mismatched nucleotides. Nearly always found are two nucleotides between the acceptor and D-arm, one nucleotide between the D-arm and anticodon arm, three to five nucleotides in the extra arm, and two to six base pairs, plus three to 12 loop nucleotides, in the D- and T- arms. Commonly the nucleotides before and after the anticodon are T and A, respectively, the two nucleotides between the acceptor and D-arm are TA, and the two most proximal nucleotides of the D-loop are As. Some tRNA genes lack the potential for base pairing in the D- or T-arms and one or both of the tRNAs for serine, in particular, almost always lack paired D-arms. There is occasionally apparent overlap of tRNA genes, without obvious explanation of how these might be processed from the presumed polycistron. This is so commonly the case for the discriminator nucleotide that it is unclear how often it is encoded by

the tRNA gene. There have been particular difficulties with annotating tRNAs that are heavily edited posttranscriptionally, such as for centipede (Lavrov, Brown, and Boore, 2000) and presumably for jumping spider (Masta and Boore, 2004) mtDNAs. With only a few exceptions (Beagley et al., 1995; Beagley, Okimoto, and Wolstenholme, 1998; Beaton, Roger, and Cavalier-Smith, 1998; Helfenbein et al., 2004), a complete set of 22 tRNA genes is found, one for each amino acid plus an additional one for each of serine and leucine since these can be encoded by two different codon families. An extra *trnM*, perhaps to separate the roles of this tRNA as initiator and elongator, has been found in the mtDNAs for each of *Mytilus* (Hoffmann, Boore, and Brown, 1992), some platyhelminths (although this was not pointed out in the manuscript; Le et al., 2000), and *Ciona* (Yokobori et al., 2003). *Ciona* and *Halocynthia* (Yokobori et al., 1999) mitochondria also have an additional tRNA for glycine that recognizes AGR codons, causing a modification of the genetic code. In practice, the search centers on finding the best possibility for each of the 22 expected tRNAs, followed by an effort to finding any others in any remaining unassigned sequences.

Figure 5 about here

Protein and rRNA encoding genes are found easily by similarity to other animal mitochondrial sequences using Blast or a similar program. It is critical to consider that there are several genetic code variations for mtDNAs (see Wolstenholme, 1992). There can be some ambiguity in assigning the precise ends of these genes without information from their transcripts. This is particularly acute for the rRNA genes, and one must generally estimate the ends within a few nucleotides based on similarity to other animals' rRNA sequences, or assume that they extend to the boundaries of adjacent genes.

For protein encoding genes, there can be ambiguity in either end. For genes with complete stop codons not overlapping the downstream gene that would produce a protein of typical length and well matched, there can be little doubt of the 3' end assignment. However, some end at abbreviated stop codons, i.e., a T at a first codon position (or a TA in the first and second codon positions), after which the transcript is enzymatically cleaved, with the stop codon completed to TAA by polyadenylation (Ojala et al., 1980). The commonality of such a T exactly juxtaposed to the first nucleotide of the downstream gene, in cases where the reading frame remains open for a great many more codons, argues strongly that this is a common mechanism. For mtDNAs of vertebrates, AGA is a stop codon, so here abbreviated stop codons can be AG in the first two codon positions that would be completed by the same mechanism.

Mitochondrial proteins initiate not only as in the "universal" code (ATG), but in some cases with alternative ATN (ATA, ATT, and ATC) and NTG (GTG, TTG, and CTG) codons. Particular animal groups may use only some of these variations and this is codified into the translation tables available at GenBank. All match at least two of the nucleotides in the CAT anticodon of the methionyl tRNA. This normally singular tRNA must deliver formyl-methionine to the initiator position during mitochondrial protein translation and methionine to internal positions, so whatever mechanism allows this discrimination presumably also enables the looser codon matching during initiation. Further, *cox1* uniquely has been found rarely to use bizarre initiation codons, including the four-member ATAA for *Drosophila* (Clary and Wolstenholme, 1985). These variations complicate the assignment of the beginnings of genes. In practice, we assume that the correct initiation site is the first eligible in frame start codon as close as possible to the extent of similarity matching and without overlapping the upstream gene.

Of course, it is necessary to do a thorough search of both strands for these genes, and to report it correctly to the sequence databases. GenBank has a very large number of errors where genes are not correctly annotated as being on the reverse complement strand.

Figure 6 about here

A significant aid in annotating mitochondrial (and chloroplast, see the chapter in this volume by Jansen et al.) genomes is called DOGMA (Wyman, Jansen, and Boore, 2004; <http://phylocluster.biosci.utexas.edu/dogma/>; Figure 6). This accepts user sequences into a password protected file, searches using COVE methods (Eddy and Durbin, 1994; Wyman and Boore, 2003) for tRNAs and presents potential secondary structures, along with the alternatives found for each, and presents aligned protein sequences. There are many user interactive features, including the ability to manually add or remove genes from the automated annotation and to designate start and stop codons by clicking on the highlighted alternatives while simultaneously viewing all annotations on a graphical display. DOGMA can extract genes or intergenic regions for subsequent analysis and can save all annotations in a table that can be read into Sequin, GenBank's submission program. Annotated sequences can also be saved in a format that can be read directly into commercial software such as MacVector (Accelrys) for subsequent analysis.

VI. SOFTWARE FOR ANALYSIS AND COMPARISONS

Once an accurate sequence is assembled and verified and the genes annotated, the analysis begins. For aspects of molecular evolution, this usually starts with a set of measurements, such as amino acid composition and hydrophilicity (etc.) of inferred proteins, nucleotide composition, codon usage patterns, and strand skew. The latter is the bias between the strands for G vs. C and T vs. A and commonly reported as G-skew (G – C)/(G + C) and T-skew (T – A)/T + A) (Perna

and Kocher, 1995), such that zero indicates perfect balance and 1 or –1 indicates total skew. [A misconception often heard is that the heavy and light strands, caused by strand skew, are somehow homologous; actually, there is no homology of heavy strands (or light strands) across animals.] Non-coding regions are often searched for potential secondary structures, which may mediate replication or transcription control, and for repeats, both direct and inverted, typically using a dot-matrix plot. These tasks are easily accomplished with MacVector (Accelrys) or other commercially available software packages (e.g., VectorNTI, GCG).

The comparison of many of these molecular features has become very difficult with so many mitochondrial sequences recently available. We are working to facilitate this by building searchable databases that will enable broad comparisons of all mtDNA sequences, complete or by individual gene, for specified sets of taxa, for features such as codon usage, nucleotide and amino acid content, gene arrangement, anticodon identities, and tRNA structures. Look for updates by following the Evolutionary Genomics/Organelles links at <http://www.jgi.doe.gov/>.

For phylogenetic analysis the relevant sequences must be aligned. Multiple sequence alignment and phylogenetic analysis of sequences is beyond the scope of this chapter, but we offer a few general comments. Due to both real variability and uncertainty of annotation, the ends of protein and rRNA encoding genes are often ambiguously aligned and are best trimmed back to a region of confidence. Alignment of genes for tRNAs and rRNAs can be guided by their potential secondary structures (Kumazawa and Nishida, 1993; Hickson et al., 1996; Macey and Verma, 1997). Regions with many gaps can contain significant alignment errors and should be eliminated or used only with caution.

VII. GENE ORDER COMPARISONS

There are many strengths to using mitochondrial gene rearrangements for phylogenetic reconstruction of animals. Nearly all animal mtDNAs contain an identical set of genes and these can potentially be reordered into an enormous number of states, so it is unlikely that reversion or convergence would occur. In many lineages these are slowly rearranging, enhancing the possibility that there will be signal at deep levels, but the finding of radically rearranged mtDNAs shows that these are not under strong selection. (Rather, it seems likely that the infrequency of rearrangement is due to the lack of recombination and the paucity of intergenic nucleotides, such that rearrangements commonly interrupt genes.) The finding of lineages that have rapidly rearranging mtDNAs does not undermine the utility of comparing gene arrangements for phylogenetic reconstruction – they are not rearranging into convergent states – but rather determines the taxonomic level at which the signal of relatedness would most likely be found. This common misunderstanding is well exemplified by the "cautionary tale" of Le et al. (2000), where a surprising number of rearrangements are described for some platyhelminths, but without demonstrating evidence that this would compromise phylogenetic reconstruction. It is specifically this lack of clocklike behavior of mitochondrial gene rearrangements that make them most useful for addressing the most difficult situation for phylogenetic reconstruction, having a short internode and a long subsequent time of divergence; there is a brief time to have accumulated signal of relatedness and a long time to erase it, where a perfectly regular rate of change is most likely to fail to reveal the signal.

Although DNA sequence comparisons continue to revolutionize our understanding of organismal relationships, they have also created an expectation that the signal of relatedness must be teased by ever more complex methods from a large body of homoplasious noise. They have revealed much of the pattern of evolution, but many branches of life that were ambiguous

in early studies have remained recalcitrant, due in part to limitations such as having only four (or 20, for amino acids) character states, alignment ambiguities, compositional bias, convergent selection, extreme rate variation, uncertainty over weighting changes of nucleotides that are paired vs. unpaired in rRNA secondary structures and, especially, difficulties with short internodes. Although gene rearrangements will define only a small number of evolutionary groups, since there may have been no rearrangements during the period of shared history or because subsequent rearrangements may have erased the signal, these can be such strong synapomorphies as to be singularly convincing. This is similar to the situation for many groups that are well accepted based on sharing very strong morphological synapomorphies, such as Tetrapoda or Mammalia.

Reconstructing phylogenetic relationships from gene arrangements can be likened to a card shuffling exercise. Imagine a stack of cards for each genome such that each card represents a single gene, which can be face up or down, analogous to transcriptional orientation. Transformation processes among these decks can be removal and replacement of one or a block of genes, with or without turning them over. (If the genome is circular, the deck must be continuous, i.e., have the bottom card also above the top card.) One can imagine constructing a network connecting these decks with branch lengths proportional to the (perhaps weighted) number of transformations. The shortest possible network would constitute a phylogenetic reconstruction.

Unfortunately, despite considerable effort (Sankoff, Cedergren, and Abel, 1990; Sankoff et al., 1992; Blanchette, Bourque, and Sankoff, 1997; Sankoff and Blanchette, 1998a,b, 1999; Blanchette, Kunisawa, and Sankoff, 1999a,b; El-Mabrouk and Sankoff, 1999; Cosner et al., 2000; Bader, Moret, and Yan, 2001; Moret et al., 2001a,b; Wang et al., 2002; Tang and Moret,

2003), no method yet exists that is completely satisfactory. Those based on distance matrices are highly subject to artifactual clustering of taxa into biologically unrealistic groupings and those that retain character information are computationally unfeasible. Further, some methods use models such as allowing only gene inversions (El-Moubrouk and Sankoff, 1999; Bader, Moret, and Yan, 2001) that are not biologically realistic for animal mtDNAs. Other work in this volume will detail these efforts, so here we will limit ourselves to the method we most commonly use, a gene adjacency matrix [equivalent to "maximum parsimony of multiple encodings" of Wang et al. (2002)], and to some general observations about gene rearrangements for phylogenetic inference.

Figure 7 about here

Simply sharing even identical gene arrangements does not necessarily indicate close evolutionary relationship. As shown in Figure 7, gene arrangements can be shared as ancestral states (i.e., sympleisiomorphies), retained from an ancestor, while other lineages have changed independently. It is important that taxa are united only when their shared arrangement can be shown to be evolutionarily derived (i.e., a synapomorphy) by comparison with that of outgroup taxa (Figure 7).

In an attempt to do this, we have developed a method (Boore et al., 1995) based on gene adjacencies as phylogenetic characters. There are 72 characters scored, each being "upstream of" or "downstream of" each of the 37 genes. Character states, then, are the 3' or 5' end of the adjacent gene, and so a matrix is filled by these, considering each animal. This matrix is then subjected to parsimony analysis (Swofford, 2001). An obvious shortcoming is that single gene boundaries are not very reliable phylogenetic characters; longer shared similarities are much stronger (e.g., Macey et al., 1997a). This matrix and analysis, however, forms an excellent

bookkeeping system that can be used to search for these longer shared arrangements by eye. In practice, the phylogenetic questions we've addressed have included gene arrangements at two extremes, either well conserved with so few gene rearrangements that synapomorphies are apparent by eye, or so rearranged that no convincing signal is likely to be recovered using any method. Table III outlines several phylogenetic relationships, some controversial and others not, well supported at this point by mitochondrial gene rearrangements.

Table III about here

What generalities can we infer from comparing animal mitochondrial gene rearrangements? First, they do not follow to any extent a molecular clock. For some lineages, rearrangements are few over hundreds of millions of years. For example, the gene arrangements of sharks and human are identical, and those of Drosophila and horseshoe crab differ only by the location of one tRNA gene. For others, rearrangements are much more rapid. For example, there are nearly no gene boundaries in common between a polyplacophoran (Boore and Brown, 1994) and bivalve (Hoffmann, Boore, and Brown, 1992) mollusk mtDNAs. Gene rearrangements are found even within a genus of amphisbaenian reptiles (Macey et al., 2004) and among closely related gastropod mollusks (Rawlings, Collins, and Bieler, 2001).

Figure 8 about here

Second, the most common mode of change can be modeled by "duplication-random loss" (Boore, 2000; Lavrov, Boore, and Brown, 2002; Figure 8). In this model, a duplication is first generated, perhaps by slipped strand mispairing, illegitimate recombination, or errors during replication (such that the termination point overruns its initiation), RNA processing, or topoisomerase activity (see Boore, 2000). The supernumerary genes can then be lost without functional consequence, which may restore the original order or lead to rearrangement. In

contrast to the mode in chloroplast genomes, inversions seem to have been rare. To the best of our knowledge, none of the algorithms in development for reconstructing phylogeny incorporate gene duplication intermediates in their modeling.

Although it seems unlikely that identical duplications followed by an identical pattern of losses would occur separately in different lineages, it is possible that homoplasious rearrangements would occur if a duplication persisted through several lineage splits, such that each of several descending lineages would inherit the identically duplicated genes. Then one could imagine that identical losses might occur in less related groups. However, the infrequency with which we've observed duplicated genes in animal mtDNAs suggests that this state is usually short lived.

Third, for vertebrates, the loss of the stem-loop structure between *trnN* and *trnC* that defines the light-strand origin of replication (see Clayton, 1992) is correlated with aberrant tRNA structures and rearrangement of tRNA genes in several lineages (Desjardins and Morais, 1990; Kumazawa and Nishida, 1995; Lee and Kocher, 1995; Macey et al., 1997a,b, 2000; Rest et al., 2003). The hypothesis has been advanced (Macey et al., 1997a,b, 2004) that the loss of this stem-loop enables the use of multiple regions, perhaps structures that are a compromise between encoding a tRNA and serving this function, for replication origins. These weaker origins, then, may lead to local duplications because of errors in aligning termination with the initiation point, which then lead to gene rearrangements through subsequent random losses.

Fourth, convergence of animal mitochondrial gene arrangements is seldom found. In a very large study, 540 partial mtDNA sequences have been determined for iguanian reptiles, with 199 sharing a derived switch in the order of *trnI* and *trnQ* (Macey et al., 1997a,b; 1998, 2000a,b; Melville et al., 2001; Schulte et al., 2002, 2003, 2004; Townsend and Larson, 2002), from which

we can infer the monophyly of the Acrodonta, and 341 sharing the ancestral condition, without homoplasy or additional changes observed. Even in cases where many of the genes are "scrambled" relative to other taxa, as is the case, for example, for a bivalve mollusk (Hoffmann, Boore, and Brown, 2000), a phthirapteran insect (Shao, Campbell, and Barker, 2001), nematodes (Okimoto et al., 1991, 1992; Keddie, Higazi, and Unnasch, 1998), and platyhelminths (Le et al., 2000; Nickisch-Rosenegk, Brown, and Boore, 2001), there is no case of these separately arriving at an identical arrangement for any of the genes. Although widely cited as an example of convergent rearrangement (Mindell, Sorenson, and Dimcheff, 1998), birds do not constitute such an example; in fact, all birds studied to date share an identical arrangement of all 37 mitochondrial genes (Figure 8; see explanation in Boore and Brown, 1998). One case has been found of convergence in protein gene rearrangement (Macey et al., 2004), a switch in gene rearrangement of the block *cob, trnT, trnP* with the block *nad6*, *trnE* in rhineurid reptiles that is identical to the rearrangement shared by birds. These genes are immediately downstream of the origin of heavy strand replication. There has been a case of convergent rearrangement of two nearest neighbor tRNA genes (*trnK* and *trnD*) in orthopteran insects (Flook, Rowell, and Gellissen, 1995) and of the same tRNA pair in hymenopteran insects (Dowton and Austin, 1999). As was pointed out by Boore and Brown (1998), these are the two types of rearrangements, i.e., rearrangements immediately downstream of an origin of replication and exchange of position of nearest neighbor tRNA genes, that should be given less weight for phylogenetic inference, since they are the most subject to duplications that might lead to rearrangement in the duplication-random loss mode.

Acknowledgements

We are grateful to lab members and guests who have worked to refine many of these bench techniques, including Ronald Bonett, David Engle, Jonathan Fong, H. Matthew Fourcade, Kevin Helfenbein, Jennifer Kuehl, Kirsten Lindstrom, Susan Masta, Jenna Morgan, Rachel Mueller, Dan Mulcahy, James Parham, Gabriela Parra, Marco Passamonti, Marcos Perez-Losada, Renfu Shao, Ernesto Recuero, Inaki Ruiz-Trillo, Wes Savage, Renfu Shao, Brian Simison, Tori Takaoka, and Yvonne Vallès. Thanks to Douda Bensasson, Jeff Froula, Allen Haim, and Stacia Wyman for work on databases and software for determining, cataloging, and comparing gene arrangements. Thanks to the National Science Foundation for support (DEB-0089624, EAR-0120646, DEB-9807100). This is LBNL-55278; part of this work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Berkeley National Laboratory, under contract No. DE-AC03-76SF00098.

Literature cited

Adams, K. L., and Palmer, J. D. (2003). Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* **29,** 380-395.

Armstrong, M. R., Blok, V. C., and Phillips, M. S. (2000). A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics* **154,** 181–192.

Arndt, A., and Smith, M. J. (1998). Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. *Mol. Biol. Evol.* **15(8),** 1009-1016.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (eds.) (2004). "Current Protocols in Molecular Biology." John Wiley & Sons, NewYork.

Bader, D. A., Moret, B. M., and Yan, M. (2001). A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. *J. Comput. Biol.* **8(5),** 483-491.

Baner, J., Nilsson, M., Mendel-Hartvig, M., and Landegren, U. (1998). Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* **26,** 5073-5078.

Beagley, C. T., MacFarlane, J. L., Pont-Kingdon, G. A., Okimoto, R., Okada, N., and Wolstenholme, D. R. (1995). Mitochondrial genomes of Anthozoa (Cnidaria), *In* "Progress in Cell Research" (F. Palmieri, ed.), Vol. 5, pp. 149-153.

Beagley, C. T, Okimoto, R., and Wolstenholme, D. R. (1998). The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* **148,** 1091-1108.

Beaton, M. J., Roger, A. J., Cavalier-Smith, T. (1998). Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum*: conserved gene order among octocorals. *J. Mol. Evol.* **47,** 697-708.

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

Boore, J. L. (2000). The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. *In* "Comparative Genomics," (D. Sankoff and J. Nadeau, eds.), Vol. 1, pp. 133-147. Kluwer Academic Publishers, Dordrecht, Netherlands.

Boore, J. L., and Brown, W. M. (1994). The complete DNA sequence of the mitochondrial genome of the Black Chiton *Katharina tunicata*. *Genetics* **138,** 423-443.

Boore, J. L., and Brown, W. M. (1998). Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.* **8,** 668-674.

Boore, J. L., and Brown, W. M. (2000). Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: Sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol. Biol. Evol.* **17(1),** 87-106.

Boore, J. L., Daehler, L. L., and Brown, W. M. (1999). Complete sequence, gene arrangement and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* ("Amphioxus"). *Mol. Biol. Evol.* **16(3),** 410-418.

Boore, J. L., and Staton, J. (2002) The mitochondrial genome of the sipunculid *Phascolopsis gouldii* supports its association with Annelida rather than Mollusca. *Mol. Biol. Evol.* **19(2),** 127-137.

Boore, J. L., Collins, T. M., Stanton, D., Daehler, L. L., and Brown, W. M. (1995). Deducing arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376,** 163-165.

Boore, J. L., Lavrov, D. V., and Brown, W. M. (1998). Gene translocation links insects and crustaceans. *Nature* **392,** 667-668.

Boore, J. L., Medina, M., and Rosenberg, L. A. (2004). Complete sequences of two highly rearranged molluscan mitochondrial genomes, those of the scaphopod *Graptacme eborea* and of the bivalve *Mytilus edulis*. *Mol. Biol. Evol.* **21(8),** 1492–1503.

Blanchette, M., Bourque, G., and Sankoff, D. (1997). Breakpoint Phylogenies. *Genome Inform. Ser. Workshop Genome Inform.* **8,** 25-34.

Blanchette, M., Kunisawa, T., and Sankoff, D. (1999a). Gene order breakpoint evidence in animal mitochondrial phylogeny. *J. Mol. Evol.* **49(2),** 193-203.

Blanchette, M., Kunisawa, T., and Sankoff, D. (1999b). Parametric genome rearrangement. *Gene* **172(1),** GC11-17.

Bridge, D., Cunningham, C. W., Schierwater, B., DeSalle, R., and Buss, L. W. (1992). Class-level relationships in the phylum Cnidaria: Evidence from mitochondrial genome structure. *Proc. Natl. Acad. Sci. USA* **89,** 8750-8753.

Brown, W. M., George, M. Jr., Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **76(4),** 1967-1971.

Budowle, B., Allard, M. W., Wilson, M. R., Chakraborty, R. (2003). Forensics and mitochondrial DNA: Applications, debates, and foundations. *Annu. Rev. Genomics Hum. Genet.* **4,** 119-141.

Cheng, S., Fockler, C., Barnes, W. M., and Higuchi, R. (1994). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91,** 5695-5699.

Clary, D. O., and Wolstenholme, D. R. (1985). The mitochondrial DNA molecule of *Drosophila yakuba:* Nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22,** 252-271.

Clayton, D. A. (1992). Transcription and replication of animal mitochondrial DNAs. *Intl. Rev. Cytol.* **141,** 217-232.

Cosner, M. E., Jansen, R. K., Moret, B. M., Raubeson, L. A., Wang, L. S., Warnow, T., and Wyman, S. (2000). A new fast heuristic for computing the breakpoint phylogeny and experimental phylogenetic analyses of real and synthetic data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **8,** 104-115.

Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11,** 1095-1099.

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., and Lasken, R. S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **99,** 5261-5266.

Desjardins, P., and Morais, R. (1990). Sequence and gene organization of the chicken mitochondrial genome. *J. Mol. Biol.* **212,** 599-634.

Detter, J. C., Jett, J. M., Lucas, S. M., Dalin, E., Arellano, A. R., Wang, M., Nelson, J. R., Chapman, J., Lou, Y., Rokhsar, D., Hawkins, T. L., and Richardson, P. M. (2002).

Isothermal strand-displacement amplification: Applications for high-throughput genomics. *Genomics* **80,** 691-698.

Dowton, M., and Austin, A. D. (1999). Evolutionary dynamics of a mitochondrial rearrangement ''hot spot'' in the Hymenoptera. *Mol. Biol. Evol.* **16(2),** 298–309.

Eddy, S. R, and Durbin, R. (1994). RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22,** 2079–2088.

El-Mabrouk, N., and Sankoff, D. (1999). On the reconstruction of ancient doubled circular genomes using minimum reversals. *Genome Inform. Ser. Workshop Genome. Inform.* **10,** 83-93.

Estaban, J. A., Salas, M., and Blanco, L. (1993). Fidelity of phi29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.* **268,** 2719-2726.

Flook, P., Rowell, H., and Gellissen, G. (1995). Homoplastic rearrangements of insect mitochondrial tRNA genes. *Naturwissenschaften* **82,** 336-337.

Fire, A., and Xu, S. Q. (1995). Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. USA* **92,** 4641-5.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotechnol.* **3,** 294-299.

Gray, M. W. (1999). Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* **9,** 678-687.

Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., Burger, G. (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* **26(4),** 865-878.

Gissi, C., Reyes, A., Pesole, G., and Saccone, C. (2000). Lineage-specific evolutionary rate in mammalian mtDNA. *Mol. Biol. Evol.* **17(7),** 1022–1031.

Green, P. (1996). http://bozeman.mbt.washington.edu/phrap.docs/phrap.html.

Helfenbein, K. G., Brown, W. M., and Boore, J. L. (2001). The complete mitochondrial genome of a lophophorate, the brachiopod *Terebratalia transversa*. *Mol. Biol. Evol.* **18(9),** 1734-1744.

Helfenbein, K. G., and Boore, J. L. (2004). The mitochondrial genome of *Phoronis architecta*—Comparisons demonstrate that phoronids are lophotrochozoan protostomes. *Mol. Biol. Evol.* **21(1),** 153-157.

Helfenbein, K. G., Fourcade, H. M., Vanjani, R. G., and Boore, J. L. (2004). The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister-group to protostomes. *Proc. Natl. Acad. Sci. USA* **101(29),** 10639-10643.

Hickson, R. E., Simon, C., Cooper, A., Spicer, G. S., Sullivan, J., and Penny, D. (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* **13,** 150-169.

Hoffmann, R. J., Boore, J. L., and Brown, W. M. (1992). A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics* **131,** 397-412.

Inoue, J. G., Miya, M., Tsukomoto, K., and Nishida, M. (2001). Mitogenomic perspective on the basal teleostean phylogeny: Resolving higher-level relationships with longer DNA sequences. *Mol. Phylogenet. Evol.* **20,** 275-285.

Ingman, M., Kaessmann, H., Pääbo, S., and Gyllensten, U. (2001). Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708-713.

Kearney, M., and Stuart, B. (2004). Repeated evolution of limblessness and digging heads in worm lizards revealed by DNA from old bones. *Proc. R. Soc. Lond.* B, in press.

Keddie, E. M., Higazi, T., and Unnasch, T. R. (1998). The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Mol. Biochem. Parasitol.* **95(1),** 111-127.

Krakauer, D. C., and Mira, A. (1999). Mitochondria and germ-cell death. *Nature* **400,** 125-126.

Kumazawa, Y., and Nishida, M. (1993). Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J. Mol. Evol.* **37,** 380-398.

Kumazawa, Y., and Nishida, M. (1995). Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. *Mol. Biol. Evol.* **12,** 759-772.

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

Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren. R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Gray, M. W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387,** 493-497.

Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33,** 351-397.

Lavrov, D. V., Brown, W. M., and Boore, J. L. (2000). A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forticatus*. *Proc. Natl. Acad. Sci. USA* **97,** 13738-13742.

Lavrov, D. V., Boore, J. L., and Brown, W. M. (2002). Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: Duplication and non-random loss. *Mol. Biol. Evol.* **19(2),** 163-169.

Lavrov, D., Brown, W. M., and Boore, J. L. (2004) Phylogenetic position of the Pentastomida and (pan)crustacean relationships. *Proc. R. Soc. Lond., B, Biol. Sci.* **271,** 537-544.

Le, T. H., Blair, D., Agatsuma, T., Humair, P. F., Campbell, N. J., Iwagami, M, Littlewood, D. T., Peacock, B., Johnston, D. A., Bartley, J., Rollinson, D., Herniou, E. A., Zarlenga, D. S., and McManus, D. P. (2000). Phylogenies inferred from mitochondrial gene orders-a cautionary tale from the parasitic flatworms. *Mol. Biol. Evol.* **17(7),** 1123-1125.

Lee, W.-J., and Kocher, T. D. (1995). Complete sequence of a Sea Lamprey (*Petromyzon marinus*) mitochondrial genome: Early establishment of the vertebrate geneome organization. *Genetics* **139,** 873-887.

Liu, D., Dubendiek, S. L., Zillman, M. A., Ryan, K., and Kool, E. T. (1996). Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases. *J. Am. Chem. Soc.* **118,** 1587-1594.

Lizardi, P. M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. C., and Ward, D. C. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19,** 225-232.

Macey, J. R., Larson, A., Ananjeva, N. B., Fang, Z., and Papenfuss, T. J. (1997a). Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol. Biol. Evol.* **14,** 91-104.

Macey, J. R., Larson, A., Ananjeva, N. B., and Papenfuss, T. J. (1997b). Evolutionary shifts in three major structural features of the mitochondrial genome among Iguanian lizards. *J. Mol. Evol.* **44,** 660-674.

Macey, J. R., Larson, A., Ananjeva, N. B., and Papenfuss, T. J. (1997c). Replication slippage may cause parallel evolution in the secondary structures of mitochondrial transfer RNAs. *Mol. Biol. Evol.* **14,** 30-39.

Macey, J. R., Schulte, J. A. II, Ananjeva, N. B., Larson, A., Rastegar-Pouyani, N., Shammakov, S. M., and Papenfuss, T. J. (1998). Phylogenetic relationships among agamid lizards of the *Laudakia caucasia* species group: Testing hypotheses of biogeographic fragmentation and an area cladogram for the Iranian Plateau. *Mol. Phylogenet. Evol.* **10,** 118-131.

Macey, J. R., Schulte, J. A. II, Kami, H. G., Ananjeva, N. B., Larson, A., and Papenfuss, T. J. (2000a). Testing alternative hypotheses of vicariance in the agamid lizard *Laudakia caucasia* in mountain ranges on the northern Iranian Plateau. *Mol. Phylogenet. Evol.* **14,** 479-483.

Macey, J. R., Schulte II, J. A., and Larson, A. (2000). Evolution and phylogenetic information content of mitochondrial genomic structural features illustrated with acrodont lizards. *Syst. Biol.* **49,** 257-277.

Macey, J. R., Schulte, J. A. II, Larson, A., Ananjeva, N. B., Wang, Y., Pethiyagoda, R., Rastegar-Pouyani, N., and Papenfuss, T. J. (2000b). Evaluating trans-Tethys migration: An example using acrodont lizard phylogenetics. *Syst. Biol.* **49,** 233-256.

Macey, J. R., Papenfuss, T. J., Kuehl, J. V., Fourcade, H. M., and Boore, J. L. (2004). Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genome sequences. *Mol. Phylogenet. Evol.*, in press.

Macey, J. R., and Verma, A. (1997). Homology in phylogenetic analysis: Alignment of transfer RNA genes and the phylogenetic position of snakes. *Mol. Phylogenet. Evol.* **4,** 272-279.

Masta, S. E., and Boore, J. L. (2004). The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. *Mol. Biol. Evol.* **21,** 893-902.

Melville,J., Schulte, J. A. II, and Larson, A. (2001). A molecular phylogenetic study of ecological diversification in the Australian lizard genus *Ctenophorus. J. Exp. Zool.* **291,** 339-353.

Mindell, D., Sorenson, M. D., and Dimcheff, D. E. (1998). Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. USA* **95,** 10693-10697.

Moret, B. M., Wang, L. S., Warnow, T., and Wyman, S. K. (2001a). New approaches for reconstructing phylogenies from gene order data. *Bioinformatics* **17 Suppl 1,** S165-173.

Moret, B. M., Wyman, S., Bader, D. A., Warnow, T., and Yan, M. (2001b). A new implementation and detailed study of breakpoint analysis. *Pac. Symp. Biocomput.* 583-594.

Nagley, P., and Wei, Y. –H. (1998). Ageing and mammalian mitochondrial genetics. *Trends Genet.* **14,** 513-517.

Nardi, F., Spinsanti, G., Boore, J. L., Carapelli, A., Dallai, R., and Frati, F. (2003) Hexapod origins, monophyletic or paraphyletic? *Science* **299,** 1887-1889.

Nickisch-Rosenegk, M. von, Brown, W. M., and Boore, J. L. (2001). Sequence and structure of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: Gene arrangement indicates that platyhelminths are derived eutrochozoans. *Mol. Biol. Evol.* **18(5),** 721-730.

Nieminen, A. L. (2003). Apoptosis and necrosis in health and disease: Role of mitochondria. *Intl. Rev. Cytol.* **224,** 29-55.

Nikaido, M., Rooney, A. P., and Okada, N. (1999). Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* **96,** 10261–10266.

Nyakaana, S., Arctander, P., and Siegismund, H. (2002). Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. *Heredity* **89(2),** 90-98.

Ojala, D., Merkel, C., Gelfand, R., and Attardi, G. (1980). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* **22,** 393-403.

Okimoto, R., Chamberlin, H. M., MacFarlane, J. L., and Wolstenholme, D. R. (1991). Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host race identification. *Nucl. Acids Res.* **19,** 1619-1626.

Okimoto, R., MacFarlane, J. L., Clary, D. O., and Wolstenholme, D. R. (1992). The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* **130(3),** 471-498.

Palumbi, S., Martin, A., Romano, S., McMillan, W. O., Stice, L., and Grabowski, G. (1991). "The Simple Fool's Guide to PCR." Version 2.0., University of Hawaii, Department of Zoology and Kewalo Marine Laboratory.

Parsons, T. J., and Coble, M. D. (2001). Increasing the forensic discrimination of mitochondrial DNA testing through the analysis of the entire mitochondrial DNA genome. *Croatian Med. J.* **42,** 304-309.

Perna, N. T., and Kocher, T. D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41,** 353-358.

Raimond, R., Marcade, I., Bouchon, D., Rigaud, T., Borry, J. –P., and Souty-Grosset, C. Organization of the large mitochondrial genome in the isopod *Armadillidium vulgare*. *Genetics* **151,** 203–210.

Rawlings, T., Collins, T., and Bieler, R. (2001). A major mitochondrial gene rearrangement among closely related species. *Mol. Biol. Evol.* **18(8),** 1604-1609.

Rawlings, T., Collins, T., and Bieler, R. (2003). Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. *Proc. Natl. Acad. Sci. USA* **100,** 15700-15705.

Rest, J. S., Ast, J. C., Austin, C. C., Waddell, P. J., Tibbetts, E. A., Hay, J. M., and Mindell, D. P. (2003). Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol. Phylogenet. Evol.* **29,** 289-297.

Sambrook, J., and Russell, D. W. (2001). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sankoff, D., and Blanchette, M., (1998a). Multiple genome rearrangement and breakpoint phylogeny. *J. Comput. Biol.* **5(3),** 555-570.

Sankoff, D., and Blanchette, M., (1998b). Phylogenetic Invariants for Metazoan Mitochondrial Genome Evolution. *Genome Inform. Ser. Workshop Genome Inform.* **9,** 22-31.

Sankoff, D., and Blanchette, M., (1999). Phylogenetic invariants for genome rearrangements. J. Comput. Biol. **6,** 431-445.

Sankoff, D., Cedergren, R., and Abel, Y. (1990). Genomic divergence through gene rearrangement. *Methods Enzymol.* **183,** 428-438.

Sankoff, D., Leduc, G., Antoine, N., Paquin, B., Lang, B. F., and Cedergren, R. (1992). Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. *Proc. Natl. Acad. Sci. USA* **89(14),** 6575-6579.

Schmitz, J., Ohme, M., and Zischler, H. (2001). SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. *Genetics* **157,** 777–784.

Schulte J. A. II, Macey, J. R., Larson, A., and Papenfuss, T. J. (1998). Molecular tests of phylogenetic taxonomies: A general procedure and example using four subfamilies of the lizard family Iguanidae. *Mol. Phylogenet. Evol.* **10,** 367-376.

Schulte II, J. A., Macey, J. R., Pethiyagoda, R., and Larson, A. (2002). Rostral horn evolution among agamid lizards of the genus *Ceratophora* endemic to Sri Lanka. *Mol. Phylogenet. Evol.* **22,** 111-117.

Schulte II, J.A., Melville, J., and Larson, A. (2003). Molecular phylogenetic evidence for ancient divergence of lizard taxa on either side of Wallace's Line. *Proc. R. Soc. Lond., B, Biol. Sci.* **270,** 597-603.

Schulte,J.A. II, Vindum, J. V. Win, H., Thin, T., Lwin, K. S., Shein, A. K., and Tun, H. (2004). Phylogenetic relationships of the genus *Ptyctolaemus* (Squamata: Agamidae), with a description of a new species from the Chin Hills of Western Myanmar. *Proc. Calif. Acad. Sci.* **55,** 222-247.

Scouras, A., and Smith, M. J. (2001). A novel mitochondrial gene order in the crinoid echinoderm *Florometra serratissima*. *Mol. Biol. Evol.* **18 (1),** 61-73.

Shao, R., Campbell, N. J. H., and Barker, S. C. (2001). Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera) *Mol. Biol. Evol.* **18(5),** 858-865.

Sorenson, M. D., Ast, J. C., Dimcheff, D. E., Yuri, T., Mindell, D. P. (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol.* **12(2),** 105-114.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98(3),** 503-17.

Swofford, D. L. (2001). "PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods)." Beta Version 4.0b8. Sinauer, Sunderland, Massachusetts.

Tang, J., and Moret, B. M. (2003). Scaling up accurate phylogenetic reconstruction from gene-order data. *Bioinformatics* **19 Suppl 1,** i305-312.

Townsend, T., and Larson, A. (2002). Molecular phylogenetics and mitochondrial genomic evolution in the Chamaeleonidae (Reptilia, Squamata). *Mol. Phylogenet. Evol.* **23,** 22-36.

Wallace, D. C. (1999). Mitochondrial diseases in man and mouse. *Science* **283,** 1482-1488.

Wang, L. S., Jansen, R. K., Moret, B. M., Raubeson, L. A., and Warnow, T. (2002). Fast phylogenetic methods for the analysis of genome rearrangement data: an empirical study. *Pac. Symp. Biocomput.* 524-535.

Wolstenholme, D. R. (1992). Animal mitochondrial DNA: Structure and evolution, in *Mitochondrial Genomes*, K. W. Jeon and D. R. Wolstenholme, eds., *Intl. Rev. Cytology* **141,** 173-216.

Wolstenholme, D. R., MacFarlane, J. L., Okimoto, R., Clary, D. O., and Wahleithner, J. A. (1987). Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc. Natl. Acad. Sci. USA* **84,** 1324-1328.

Wu, W., Schmidt, T. R., Goodman, M., and Grossman, L. I. (2000). Molecular evolution of cytochrome c oxidase subunit I in primates: Is there coevolution between mitochondrial and nuclear genomes? *Mol. Phylogenet. Evol.* **17(2),** 294-304.

Wyman, S. K., and Boore, J. L. (2003). Annotating animal mitochondrial tRNAs: An experimental evaluation of four methods. pp. 44-46 *in* Proc. European Conf. Computational Biol. (ECCB).

Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, in press.

Yokobori, S. –I., Rakuya, U., Feldmaier-Fuchs, G., Pääbo, S., Ueshima, R., Kondow, A., Nishikawa, K., and Watanabe, K. (1999). Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). *Genetics* **153,** 1851-1862.

Yokobori, S., Watanabe, Y., and Oshima, T. (2003). Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): Comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. *J. Mol. Evol.* **57,** 574-587.

Yost, H. J., Phillips, C. R., Boore, J. L., Bertman, J., Whalen, B., and Danilchik, M. V. (1995). Relocation of mitochondrial RNA to the prospective dorsal midline during *Xenopus* embryogenesis. *Dev. Biol.* **170,** 83-90.

Zardoya, R., and Meyer, A. (2001). On the origin of and phylogenetic relationships among living amphibians. *Proc. Natl. Acad. Sci. USA* **98(13),** 7380-7383.

Table I. Web resources for further information on mtDNA and mitochondrial systems.

URL	Description
http://www.jgi.doe.gov/programs/comparative/top_level/organelles.html	Organelle Genomics at DOE Joint Genome Institute
http://phylocluster.biosci.utexas.edu/dogma/	Tools for gene annotation of mtDNAs and cpDNAs
http://megasun.bch.umontreal.ca/ogmpproj.html	Organelle Genome Megasequencing Project
http://www.ncbi.nlm.nih.gov/genomes/organelles/organelles.html	Organelle genome resources at GenBank
http://www.mitomap.org/	Mitomap: Human mitochondria and disease
http://www-lecb.ncifcrf.gov/mitoDat/	Nuclear genes with mitochondrial products database

Table II. Primers in common use in our laboratory that are designed to match conserved regions and have been broadly successful across the Metazoa.

Gene	Primer	Sequence	Reference
rrnS	12SaL	AAACTGGGATTAGATACC CCACTAT	Palumbi et al.,
	12SaiL	AAACTAGGATTAGATACCCTATTAT	1991
	12SbH	GAGGGTGACGGGCGGTGTGT	
rrnL	16SarL	CGCCTGTTTATCAAAAACAT	
	16SbL	ACGTGATCTGAGTTCAGACCGG	
	16SaH	ATGTTTTTGATAAACAGGCG	
	16SbrH	CCGGTCTGAACTCAGATCACGT	
	16S1148H	ATTAYGCTACCTTWGCACRGTCARRRTACYGCGG	This publication
	16S1148L	CCGCRGTAYYYTGACYGTGCWAAGGTAGCRTAAT	
cox1	LCO1490	GGTCAACAAATCATAAAGATATTGG	Folmer et al.,
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	1994
cob	cobF424	GGWTAYGTWYTWCCWTGRGGWCARAT	Boore and
	cobR876	GCRTAWGCRAAWARRAARTAYCAYTCWGG	Brown, 2000
cox3	cox3F	TGGTGGCGAGATGTKKTNCGNGA	
	cox3R	ACWACGTCKACGAAGTGTCARTATCA	

Table III. Some of the phylogenetic relationships that are well supported by shared-derived rearrangements of mitochondrial genes (even though monophyly of some groups – crocodilians, birds, vertebrates, and echinoderms – were not controversial), or well supported by comparisons of complete, or nearly complete mtDNA sequences for cases where other studies had been equivocal or contradictory. In a few of these references, the phylogenetic conclusions were not emphasized, but the data presented can easily be interpreted to support these conclusions. See Boore and Brown (1998) and Boore (1999) for a discussion of some of these earlier results.

Relationship supported	**Reference**
Based on gene rearrangements:	
Cestode platyhelminthes are within the Eutrochozoa	Nickisch-Rosenegk, Brown, and Boore (2001)
Opisthobranchia unites with Pulmonata within Gastropoda	Kurabayashi and Ueshima (2000)
Annelida is more closely related to Mollusca than to Arthropoda	Boore and Brown (2000)
Onychophora does not group with Hexapoda to form the "Uniramia" and	Boore et al. (1995); Lavrov, Brown, and Boore
Arthropoda and Pancrustacea are each monophyletic	(2003)
Myriapoda is outside Pancrustacea, not the sister group to insects	Boore, Lavrov, and Brown (1998)
Pentasomida is not a phylum, but rather a type of crustacean	Lavrov, Brown, and Boore (2003)
Phoronida is part of the Eutrochozoa	Helfenbein and Boore (2004)
Monophyly of Crocodylidae	Kumazawa and Nishida (1995)
Monophyly of Aves	Mindell, Sorenson, and Dimcheff (1998)
Monophyly of Vertebrata to the exclusion of Cephalochordata	Boore, Daehler, and Brown (1999)
Monophyly of Echinodermata	Scouras and Smith (2001)
Hydrozoa, Scyphozoa, and Cubozoa group within Cnidaria	Bridge et al. (1992) (shared mtDNA structure)
Monophyly of acrodont lizards	Macey et al., (1997a,b, 1998, 2000a,b); Melville et al. (2001); Schulte et al. (2002, 2003, 2004); Townsend and Larson (2002)
Monophyly of the lizard family Iguanidae	Macey et al. (1997b); Schulte et al. (1998)
Bipeded amphisbaenians are derived relative to limbless groups	Macey et al. (2004)

Based on mtDNA sequence comparisons:	
Chaetognatha is basal to protostomes	Helfenbein et al. (2004)
Pogonophora is not a phylum, but a member of Annelida	Boore and Brown (2000)
Sipuncula is more closely related to Annelida than to Mollusca	Boore and Staton (2002)
Collembolida is basal to the Pancrustacea	Nardi et al. (2003)
Brachiopoda is part of the Eutrochozoa	Helfenbein and Boore (2004)
Salamanders group with frogs to the exclusion of caecilians	Zardoya and Meyer (2001)

Protocol 1: Physical isolation of mtDNA

Solutions (and storage conditions)

Homogenization buffer (4° C):

210	mM	Mannitol (adds viscosity for more gentle cell lysis)
70	mM	Sucrose (enhances mitochondrial integrity)
50	mM	Tris HCl, pH 7.5 (buffers pH changes)
3	mM	$CaCl_2$ (protects nucleoprotein complexes for nuclear membrane integrity)

(This cannot be autoclaved, but can be sterilized by passage through a 0.22 micron filter)

500 mM EDTA (4° C)
TE: 10 mM Tris, 1 mM EDTA, pH 8.0 (room temperature)
2% bleach solution (room temperature)
For a sucrose step gradient: 1 M sucrose in TE and 1.5 M sucrose in TE (4° C)
10% SDS: 10 grams SDS, water to 100 ml (room temperature)
Saturated CsCl is 7M in water
Propidium iodide (PI) solution: 2 mg/ml in water

CsCl/PI solution, density 1.57g/ml (makes 100 ml, room temperature):

74.2	g	CsCl
10.0	ml	10 x TE
54.9	ml	water
16.6	ml	PI solution (@2 mg/ml)

For a "velocitization", make this solution at two other densities also: For 1.4 g/ml, modify to 53.3 g of CsCl and 71.7 ml water. For 1.7 g/ml, modify to 93.3 g of CsCl and 61.7 ml water.
Water-saturated butanol: Mix equal portions, shake, let the layers form. (Butanol is on top.)

Procedure:

This is for small amounts of tissue, less than about 0.3 grams. Scale up for larger amounts.

1. Homogenize tissue in 3 ml of cold homogenization buffer using three, 5 sec. strokes of the tissumizer (Tekmar). A dounce homogenizer can be used instead. This may need optimizing.
2. Add 600 µl of 500 mM EDTA. Keep on ice. (Clean the tissumizer between samples by running for 15 sec in cold 2% bleach solution, then rinsing with three iterations of cold, high purity water.)
3. Pellet nuclei at 1200 X G for 5 minutes at 4° C. Remove the supernatant into a fresh, sterile tube with a pipette. If working with a large sample, this step can be repeated.
4. Pellet mitochondria at 23,000 x G for 25 minutes at 4° C. Pour off and discard the supernatant. Allow the tube to drain for a few minutes, then put on ice.

———————

Some would add a sucrose step gradient at this point to separate mitochondria from other cellular constituents, but this is not generally necessary unless the amount of tissue was very large.

A. Prepare step gradients by placing 10 ml of 1 M sucrose in TE in swinging bucket centrifuge tube. Using a pipette, underlayer this with 8 ml of 1.5 M sucrose in TE.
B. Resuspend pellet in 20 ml homogenization buffer and carefully layer on top of the gradient.
C. Centrifuge at 33,000 X G for 60 minutes at 4° C in a swinging bucket rotor. (Microsomes collect at the top of the 1 M sucrose, mitochondria at the interface of 1.5 M and 1 M sucrose, and nuclei at the bottom.)
D. Aspirate off the excess at the top, then collect the interface layer between the 1.5 M and 1 M sucrose with a short pipette, trying for minimum volume (less than 5 ml).
E. Add about 4 volumes of cold TE to this collected fraction to reduce the sucrose concentration.
F. Pellet mitochondria at 23,000 X G for 25 minutes at 4° C, then pour off and discard the supernatant. Allow the tube to drain for a few minutes, then put on ice.

———————

5. Resuspend pellet in 1.6 ml TE with vigorous vortexing at room temperature.

6. Lyse the mitochondria by adding 0.4 ml of 10% SDS and mixing gently, then let stand for 10 minutes at room temperature.
7. Add 0.33 ml saturated CsCl, mix, place on ice for at least 15 minutes. (Can remain at 4° C overnight.)
8. Pellet the mitochondrial membrane debris at 17,000 X G for 10 minutes at 4° C, then collect the supernatant into a sterile culture tube.
9. Add 0.5 ml of PI solution. Measure the volume and add solid CsCl according to the following schedule, interpolating as necessary: 2.2 ml / 2.04 g; 2.3 ml / 2.13 g; 2.4 ml / 2.22 g; 2.5 ml / 2.32 g.
10. Invert several times to dissolve the CsCl. Check the density by weighing 1 ml to ensure it is 1.57 +/- 0.01 g. Adjust using solid CsCl or water to 1.57 g/ml. Top off the tube with the CsCl/PI stock solution of density 1.57 g/ ml. Mix. Balance the tubes in pairs to within 0.02 g.
11. Ultracentrifuge to separate linear and circular DNAs. The vertical rotor spins harder than the swinging bucket rotor, so the gradient sets up faster; however, the shorter gradient produces a less pure product. The possibilities are, in order of speed vs. purity: 1. single spin in the vertical rotor; 2. single spin in the swinging bucket rotor; 3. spin in the vertical rotor followed by the swinging bucket rotor; 4. spin in the vertical rotor, a “velocitization” removing small DNA fragments, then spin in the swinging bucket rotor. Parameters for spinning in the vertical rotor: 55,000 RPM for 24 hours at 20° C. Parameters for spinning in the swinging bucket rotor: 36,000 RPM for 36-48 hours at 20° C. Consult the machine instructions for details. For the swinging bucket rotor, layer light mineral oil on top to within 3 mm of the top balancing them using the mineral oil.

    A velocitization is sometimes used between two density gradient ultracentrifugations to remove small DNA fragments: A) Measure the volume of the sample (V) and calculate X = 3.8 – (2V + 0.7). Add X ml of a 1.4 g/ml CsCl /PI solution to a tube for the swinging bucket rotor. B) Underlayer 0.7 ml of a 1.7 g/ml CsCl/PI solution by placing a Pasteur pipette tip against the side of the tube, gently releasing a drop, then sliding the pipette down to the bottom of the tube. Slowly withdraw the pipette as the last bit of liquid is delivered, keeping it against the side of the tube. C) Add V ml of TE to the sample and layer this on top of the gradient. Top off with light mineral oil and balance. D) Swinging bucket rotor run parameters: 45,000 RPM for 3 hours and 30 minutes at 20° C. Brake should be on while the machine is accelerating, but must be turned off before the spin comes down. E) Collect by puncturing tube and collecting the bottom 1.4 ml without using UV light. The intact mtDNA will have been at the interface of the two CsCl density solutions.
12. After each ultracentrafugation (except a velocitization), collect sample using a needle or a tube dripping apparatus and a UV light. There will be as many as four florescent bands. Against the bottom of the tube is RNA. At the top are carbohydrates. Nuclear DNA is about the middle of the tube. About 5 mm below the nuclear will be the mtDNA. Even if not visible, collect this region of the gradient. See details in Sambrook and Russell (2001). If another ultracentrifugation is desired and the mtDNA band is not visible, consider collecting a small amount of the nuclear band to use as a trace.
13. Add about 500 µl of water saturated butanol to the sample, shake, then let the layers form. The butanol will absorb much of the propidium iodide. Remove this with a Pasteur pipette and discard, then repeat about 5 times, until the sample appears clear, then one more time. Check for florescence with UV light. If volume reduction is desired, extract with straight butanol, which will adsorb water.
14. Remove the CsCl either by using ultrafiltration (e.g., using ultrafreee spin columns; Millipore) or dialysis. If dialyzing, use a large volume of diasylate (TE or water), at least 200 ml, with 4 buffer changes, over about 24 hours, mixing gently on a stir plate.

Protocol 2: Mapping mtDNA by radiolabeling fragments

Perform restriction enzyme digests of purified mtDNA as desired, each in 10 µl total volume.
Inactivate the enzyme by heating. In most cases, 10 minutes at 70º C is sufficient. Spin to bottom of tube.
For calculations below, n = number of digests + 2 for standards + 1 for pipetting error.
This is acceptable for up to 20 reactions; scale up for more.
In addition to filling in at 5' overhangs, the polymerase has an exonuclease activity that erodes 3' overhanging or blunt ends, which are filled back in, and will also translate from any nicks in the DNA.
The acrylamide gel allows very accurate sizing of fragments from about 20 to 1,500 base pairs, while the agarose gel best sizes larger fragments. Choose size standards with this in mind.
Carefully follow good procedures for safely handling radioactive materials.

Reaction mix:
1.5n µl reaction buffer
2 µl equal mixture of alpha labeled radionucleotides, 300 µCi/ml
1 µl Klenow polymerase
Add water to bring volume to 5n µl

Add 5 µl of reaction mix to each restriction enzyme digest, including 2 for size standards.
Incubate for 30 minutes, 5' overhang digests on ice, 3' overhang or blunt end digests at 37º C.

Divide this evenly into two fractions, add 2 µl of gel loading buffer ("bluecrose"), and load onto each of an acrylamide (3.5 to 6%) and agarose (0.8 to 1.2%) gel, using TBE, normally in vertical gel rigs. Load bluecrose into empty lanes.

After gel is run, remove to 3 MM paper, cover with plastic wrap, and dry in a vacuum gel dryer.
Expose to X-ray film, nominally for 24 hours, although this can be repeated for shorter or longer times.
The result looks like a negative of an ethidium bromide stained gel.

10X reaction buffer: 60 m<u>M</u> KCl, 100 m<u>M</u> Tris (pH 7.2), 100 m<u>M</u> $MgCl_2$, 70 m<u>M</u> beta-mercaptoethanol

Protocol 3: RCA with specific primers

Denaturing reaction:

4 µl	DNA template
4 µl	10 µM primer mix (at least one for each strand)
2 µl	5X denaturing buffer (from TempliPhi kit; Amersham)
10 µl	Final volume

Denature at 95° C for 1 minute. Cool to 34° C.

Mix:

2 µl	10X reaction buffer
5 µl	High quality water
2 µl	2 mM dNTPs
1 µl	Phi29 enzyme
10 µl	Final volume

Add this mix to the denatured sample at 34° C for 15 hours, then heat to 65° C, 15 minutes. Store at 4° C.

10X reaction buffer: 75 mM NaCl, 60 mM $MgCl_2$, 1 mM DTT, 0.01% Tween-20, 25 mM Tris pH 8.1 (pH adjusted with boric acid rather than HCl)

Figure legends

Figure 1: Phylogenetic distribution, by phylum, of the complete mtDNA sequences in GenBank as of June 2004, with a list of those phyla remaining unrepresented.

Figure 2: Products of physical purification and of a rolling circle amplification reaction. Lanes are as follows: 1. KB ladder standard; 2. mtDNA purified from *Xenopus* eggs using the technique in Protocol 1; 3. This same preparation treated with RNAse; 4. KB ladder standard; 5. RCA product using the technique in Protocol 3 from 1/100 of the material in lane 2; 6. This RCA product digested with *Spe*I; 7. The purified mtDNA shown in lane 2 digested with *Spe*I.

Figure 3: Long PCR amplification products on agarose gels stained with ethidium bromide. Multibanded standards in each case are KB ladder. The first amplification is the nearly complete mtDNA (14,465 nts) of the scaphopod mollusk *Graptacme eborea* (Boore, Medina, and Rosenberg, 2004). Although this was useful, we much more commonly amplify mtDNAs in two or three generously overlapping fragments, as shown in the remaining panels.

Figure 4: Breaking DNA into random fragments with a HydroShear device (GeneMachines). (A) Long strands of DNA are loaded into the device, then driven under pressure repeatedly through a narrow aperture, shearing them to a size specified by the pressure applied. (B) Agarose gel stained with ethidium bromide showing a typical result. The standard in lane 1 has bands at 1, 2, 3, 4, 6 and 10 kb. Note the narrow distribution of product size.

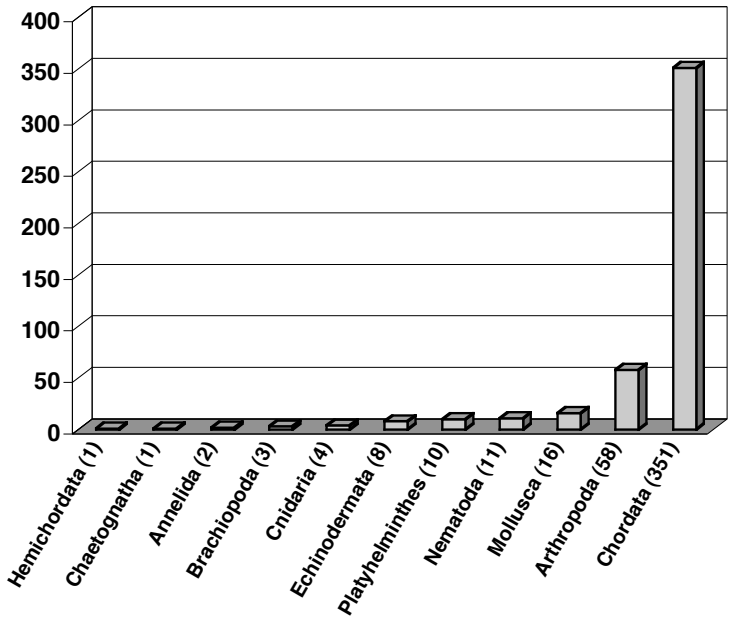
Figure 5: An arbitrarily selected valine tRNA in schematic cloverleaf form with each section labeled.

Figure 6: One screen from DOGMA (Dual Organellar GenoMe Annotator) available at <http://phylocluster.biosci.utexas.edu/dogma/> (Wyman, Jansen, and Boore, 2004) showing the folding of an identified tRNA gene along with alternative possibilities for this same gene elsewhere in the mtDNA. Shown below is part of the graphical overview of the gene annotation produced by the software.

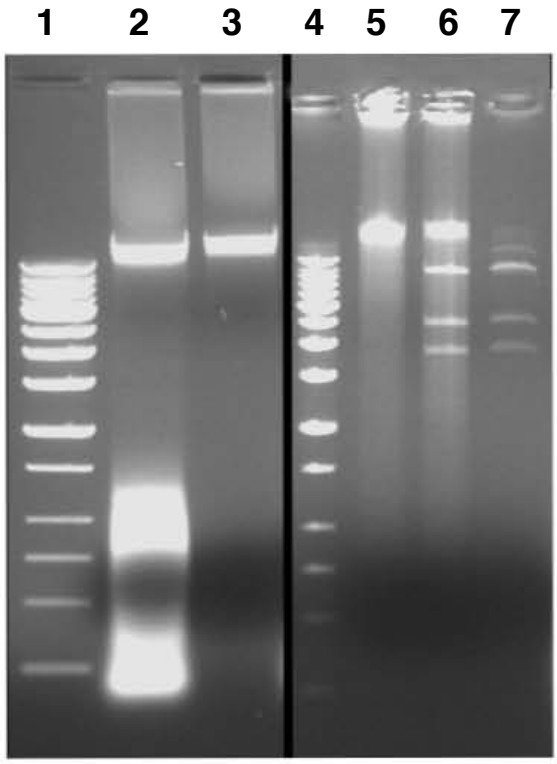
Figure 7: Having identical gene arrangements does not necessarily indicate close relationship. Lower case letters label genes in this hypothetical example. The gene arrangement for Taxa 2 and 3 are identical, but trees A, B, and C are equally parsimonious. Two changes, indicated by asterisks, are required for each tree. (This also shows the primary shortcoming of the gene adjacency method, and the reason why single gene boundaries are weak phylogenetic characters; independent movements of gene "c" creates gene boundary "b, d" in parallel.) A valid phylogenetic reconstruction requires the situation shown in tree D, where (at least) two taxa are in one arrangement, with (at least) two in an alternative arrangement. This tree requires only one change, whereas any other relationship among these taxa would be less parsimonious.

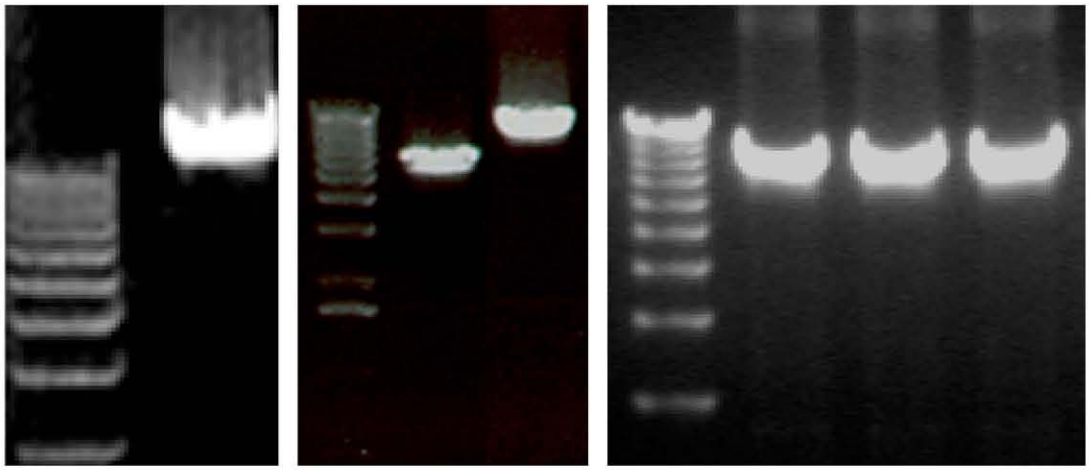
Figure 8: Illustration of the duplication-random loss model of gene rearrangements (see Boore, 2000). Genes for tRNAs are abbreviated by the one-letter code for the corresponding amino acid. No annotation is made here for transcriptional orientation. (A) Eutherian and marsupial mammals differ in the arrangement of several genes (see Boore, 1999). A hypothetical

intermediate includes the duplication of four of the genes. Losses of the supernumerary genes, then, would be expected to have little functional consequence. Some patterns of loss could potentially have restored the original order, but others could have led to the rearrangement in marsupials. (B) A model of tandem duplication of a two gene block, followed by random losses of supernumerary genes (Macey et al., 1997a, 1998), could have led to the rearrangement found in acrodont lizards (references in Table III). Even after duplication, three of the four possible patterns of loss would have led to restoration of the original order. A similar case has been found for *trnP* and *trnT* in the amphisbaenian genus *Bipes* (Macey et al., 2004). (C) An intermediate in this type of process seems to have been found in the sea cucumber *Cucumaria* (Arndt and Smith, 1998). Here, one region of the mtDNA matches partially the arrangement easily inferred to be ancestral for this echinoderm class, with blocks of unassignable nucleotides (indicated by numerals) in place of genes missing from the array. At another location in the mtDNA are the "missing" genes, again separated by blocks of unassignable nucleotides. One might infer that this stems from an ancestral duplication and movement of the tRNA gene cluster, followed by random gene degradation leading to this arrangement, with the unassigned blocks being unrecognizable vestiges of tRNA genes. (D) The rearrangement at the base of birds was a switch in order of the blocks *cob*, *trnT*, *trnP* and *nad6*, *trnE* to generate the order shown for the "hypothetical intermediate", except here we speculate that there was an additional "NC" (non-coding region) between *trnT* and *trnP*. Random losses of these two non-coding regions could generate the conditions found for modern birds. Although Mindell, Sorenson, and Dimcheff (1998) interpret this as a gene rearrangement (and do not view the NC between *trnE* and *trnF* in order C as homologous to the corresponding region in order B, even though there is significant sequence similarity), it is really the position of these non-coding regions that vary.

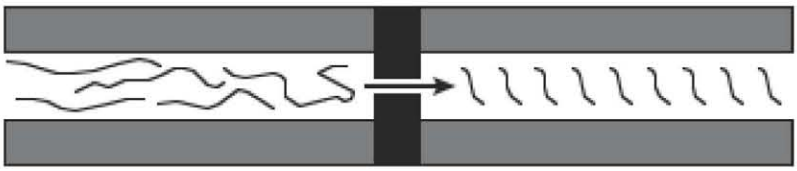


None for the phyla:
Porifera
Ctenophora
Kinorhyncha
Loricifera
Gnathostomula
Gastrotricha
Rotifera
Acanthocephala
Nemertea
Nematomorpha
Bryozoa
Vestimentifera
Echiura
Tardigrada
Priapula

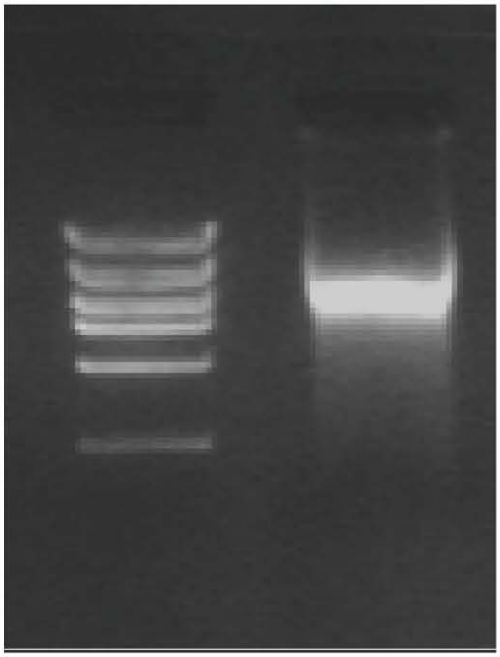

1
2
3
4
5
6
7

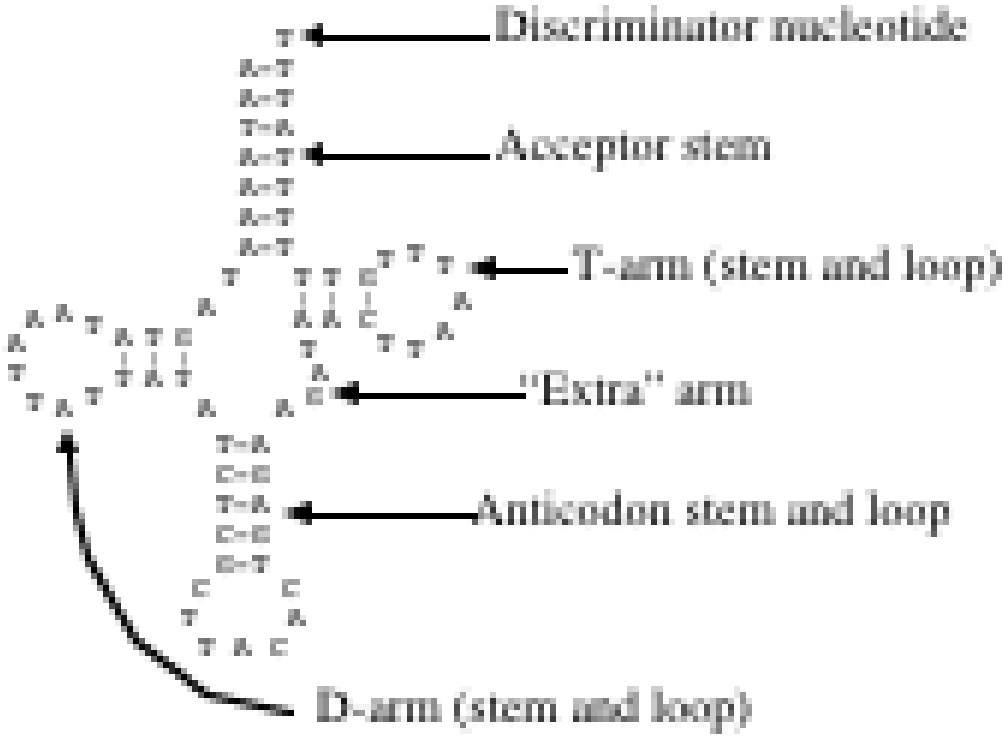


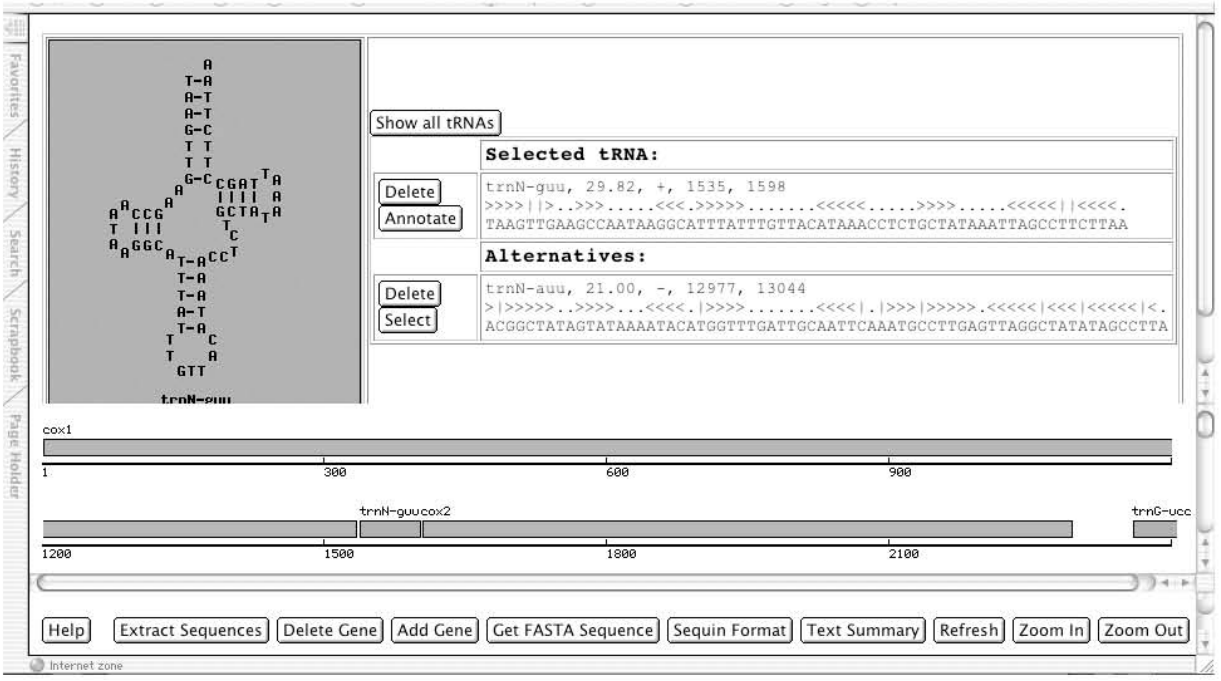
A

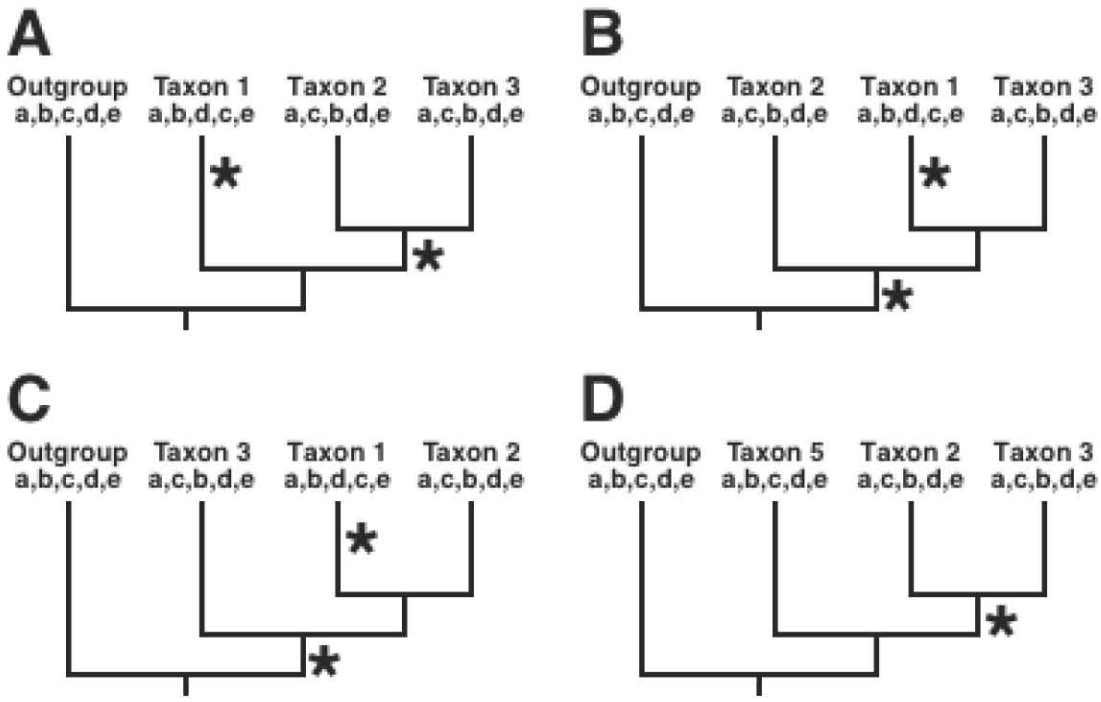


B




Discriminator nucleotide
Acceptor stem
T-arm (stem and loop)
"Extra" arm
Anticodon stem and loop
D-arm (stem and loop)


Favorites
History
Search
Scrapbook
Page Holder
trnN-guu
Show all tRNAs
Selected tRNA:
Delete
Annotate
trnN-guu, 29.82, +, 1535, 1598
>>>>||>..>>>.....<<<.>>>>>.......<<<<<.....>>>>.....<<<<<||<<<<.
TAAGTTGAAGCCAATAAGGCATTTATTTGTTACATAAACCTCTGCTATAAATTAGCCTTCTTAA
Alternatives:
Delete
Select
trnN-auu, 21.00, -, 12977, 13044
>|>>>>>..>>>>...<<<<.|>>>>.......<<<<|.|>>>|>>>>>.<<<<<|<<<|<<<<<|<.
ACGGCTATAGTATAAAATACATGGTTTGATTGCAATTCAAATGCCTTGAGTTAGGCTATATAGCCTTA
cox1
1
300
600
900
trnN-guucox2
trnG-ucc
1200
1500
1800
2100
Help
Extract Sequences
Delete Gene
Add Gene
Get FASTA Sequence
Sequin Format
Text Summary
Refresh
Zoom In
Zoom Out
Internet zone


A
Outgroup
a,b,c,d,e
Taxon 1
a,b,d,c,e
Taxon 2
a,c,b,d,e
Taxon 3
a,c,b,d,e
B
Outgroup
a,b,c,d,e
Taxon 2
a,c,b,d,e
Taxon 1
a,b,d,c,e
Taxon 3
a,c,b,d,e
C
Outgroup
a,b,c,d,e
Taxon 3
a,c,b,d,e
Taxon 1
a,b,d,c,e
Taxon 2
a,c,b,d,e
D
Outgroup
a,b,c,d,e
Taxon 5
a,b,c,d,e
Taxon 2
a,c,b,d,e
Taxon 3
a,c,b,d,e

A

Eutherian arrangement	nad2, W, A, N, C, Y, cox1
Hypothetical duplication	nad2, W, A, N, C, W, A, N, C, Y, cox1
To restore original order	nad2, W, A, -, -, -, -, N, C, Y, cox1
For marsupial order	nad2, -, A, -, C, W, -, N, -, Y, cox1
Marsupial arrangement	nad2, A, C, W, N, Y, cox1

B

C

Ancestral echinoderm	rrnS, E, T, 120, P, Q, N, L, A, W, C, V, M, D, Y, G, L, nad1
Cucumaria region 1	rrnS, 64, T, 459, Q, 98, A, 30, C, 60, M, D, Y, G, L, nad1
Cucumaria region 2	1030, E, 410, P, 20, N, L, 63, W, 58, V, 79, nad4L

D

Hypothetical ancestor	nad5, cob, T, NC, P, nad6, E, NC, F, rrnS
Bird order B	nad5, cob, T, --, P, nad6, E, NC, F, rrnS
Bird order C	nad5, cob, T, NC, P, nad6, E, NC, F, rrnS