Auchenorrhyncha PEET Workshop
C. Simon, November 2008

Practical Considerations, Organization, and Data Checks for Molecular Systematics Students

At every step along the way in your field and laboratory work, it is important to have checks. This document provides suggestions for molecular systematics studies based on experiences from the Simon lab. What follows is a series of suggestions in a checklist form. Although many of the suggestions are self evident, they are useful for new students entering a lab. Each lab deals with collection and databasing in a different way, but this document can be easily modified to fit your current protocols.

1) Before you begin a project, make an effort to contact all researchers worldwide who are currently working on your taxon. Introduce yourself; exchange ideas and research outlines.

2) In the specimen collecting stage:

Before you go into the field, review your project proposal to familiarize yourself with the experimental design (sampling scheme), goals and timeline.

A) Obtain collecting permits well before planned field trips. These can often take many months to obtain. Often National Parks Permits are needed in addition to Provincial level permits, City Permits, Board of Water Supply, etc. Increasingly, native peoples must be involved in a consultation process. Collecting on private land requires contacting the property owner in advance. If working in a foreign country collaborate with local researchers.

The Convention on Biological Diversity is relevant to international collaborations: http://www.biodiv.org/convention/articles.asp "The objectives of this Convention are the conservation of biological diversity, the sustainable use of its components and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources, including by appropriate access to genetic resources and by appropriate transfer of relevant technologies, taking into account all rights over those resources and to technologies, and by appropriate funding."

B) Investigate the ecology, geography, biogeographic boundaries, areas of endemism, and climate of the areas to be visited (This helps with packing as well as with collecting and data interpretation). Collect local field guides to plants and animals and use them.

C) Create a checklist of field gear in the back of your field notebook. Forgotten equipment can cause delays and data loss. Excessive gear can also cause problems. Consult with experienced colleagues for ideas. Gear can often be purchased in country to avoid airline excess baggage charges but plan time for outfitting the expedition upon arrival in a country (avoid arriving on a day or at a time that shops are closed).

Bring sufficient collection supplies such as presses for plants, pinning boxes for insects, leak proof vials such as scintillation vial (either cone lid or other) or screw-top Nunc cryotubes (if no money try parafilm to seal cheaper vials), and storage buffers/agents such as: dryrite (silica gel), the CTAB gel, FAA (for cytology and flower parts), and/paraldehyde buffer (for electron microscopy later). Be sure to take along a cooler to keep specimens cold on wet or dry ice; this is important even for alcohol specimens especially for nuclear genes. Consider preserving some specimens in "RNA later," a buffer that preserves RNA for cDNA extractions (useful for hard-to-sequence nuclear genes that may have long introns; however, see notes on alcohol/refrigeration by Jerry Regier 2005 in his recent web publication www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf). Your field gear should include copies of any important taxonomic publications/illustrations of the organisms under study (library search prior to the field trip).

D) Sample throughout a species range to determine species limits. What you think is one species could be two or three. (Examples from the NZ cicadas and Hawaiian damselflies). Plan to conduct multiple field seasons: sample widely and quickly the first season to establish potential patterns. Analyze the data during the off-season and then plan subsequent field seasons based on this knowledge.

E) When driving a planned route, remember that the route may change depending on road conditions and weather conditions, so don't assume that you will return soon to any given location in the same field season (collect what you need that year, the first time you visit a site if possible).

F) Keep detailed field notes (in legible handwriting) when collecting specimens including a daily log of activities, names of participants, exact locations of specimens (GPS and a written description of how to get to the location using permanent landmarks). GPS can be recorded as latitude-longitude or UTM; decide on a standard or record both. Record elevation. Record the locality code. Never leave the GPS unit on the roof or hood of your car! ; -) After the first field season, transfer the data from field notebooks to the database, export the data table to Arcview to plot on maps, export the maps from Arcview as EPS files and annotate/print using Illustrator. Create a book of maps and locality codes/locations that can be used as a reference in the field. Carry topographical atlas/maps with you. Your GPS can be used with the maps to help you if you are lost.

Label specimens on site and take notes about the biology (behavior, natural history) of the organism and its habitat/vegetation, temperature, weather conditions, other species present, etc. For aquatics, turbidity, oxygen content, salinity, conductivity, etc. are often recorded. Do this on site not later in the day (although you could fill in details later if daylight time is short). Do not rely on memory! On a related topic, if your organism has an obvious sexual signaling system, record the signals of each specimen collected (or as many as possible) and label it with the same code as the specimen (see below). Some parts of an organism maybe preserved differently for later morphological or chromosomal studies.

G) Photograph the site from several directions. Photograph specimens (every specimen if color patterns are important). A digital camera photo is an important supplement to the voucher specimen. Color cards standards can be purchased (or printed out all at one time for standardization) and placed next to specimen for photographing colored specimens if necessary. Photos can be downloaded in the field onto a portable laptop computer (well worth

the investment) and used later in web keys and web sites as well as in journal publications. Label downloaded photos with the same locality code as the specimens (see H below).

H) Include standardized specimen labels with each sample showing country, province, exact location, full date and collectors. (See Global Biodiversity Information Facility website <http://www.gbif.org/>. GBIF's mission is facilitation of the dissemination of digitized biodiversity data for global users; see also the international "Darwin Core 1, 2, and 3" <http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome>.

Include a species name label if known. Develop in advance a concise, preferably mnemonic, coding system for individual specimens or localities that can be carried through collection, extraction, cleaning, storage, PCR, sequencing, and databasing. Create a written description of this procedure and keep it in the front or back of your field and lab notebooks. This system should be understandable to a naïve reader. Example from our laboratory: 04.NZ.OL.LHA.02 stands for: 2004, New Zealand, Otago Lakes District, State Highway 6 at Lake Hawea access rd., ca. 11km N. of Lake Hawea dam, 1km past Lake Hawea lookout, specimen number 2. This code is entered into the data base along with the species identity, exact date, method of preservation, latitude, longitude, altitude, a written description of the location, and plus any notes; this number is cross referenced in the field notebook. The two-letter codes for each district for New Zealand was developed by Crosby, Dugdale and Watt (1998. NZ J. Zool 25:175-183). Districts for other countries can be found at: www.statoids.com or from the relevant country authorities.

J) Write all labels in pencil or permanent pen or printer ink (note: some permanent inks dissolve in alcohol and even if your specimen is not stored in alcohol now, it may be in the future or it may be stored near specimens in alcohol that may leak, especially during pressure changes in airplanes). Laser-printed or photocopied ink can lift off the paper eventually. Toasting labels in a microwave may help set. Pencil is best. See attached notes on permanent labels.

K) Include a label inside a vial as well as on the outside. Outside labels are easily damaged especially when written on smooth plastic. Water-soluble labels on the outside of vials can disappear with condensation after removal from a freezer! Alcohol can leach from vials and remove permanent ink labels. Alcohol can alter the bonding of manufacturers labels such that they rub off.

L) Tissues for DNA analysis in the field can be stored in liquid nitrogen (very difficult to transport by air), dry ice (easier to transport small amounts by air), alcohol (also difficult to transport by airCone lidded scintillation vials or cryovials with o-rings seem to hold alcohol well, as do screw-top Nunc tubes; otherwise use parafilm). Alcohol may be tipped off for air transport and refilled at the other end. Specimen tissues dilute alcohol so replacing alcohol once or twice is a good idea. For large volumes of samples, special Dangerous Goods Shippers must be used. Other storage options include silica gel drying agent or in special high salt/high EDTA buffers or CTAB. Specimens should always be kept cold. The ancient DNA folks say that water and heat are the biggest enemies of DNA. Decide on the number of replicate tissue samples needed for depositing in museums. For herps in the Patagonia project, they create three tissue duplicates (in ethanol) and two stay in the host country. For fishes, a single blood sample is taken and divided later in the lab.

M) For large animals, the voucher specimen may be preserved separately and in a different manner from the tissue samples (e.g. for future morphological studies). Collect museum voucher specimens at each locality to preserve all important characters including color (dry specimens as quickly as possible to retain colors; if using formalin to preserve color be sure to include a note on the label because formalin makes extraction extremely difficult). Label the voucher with the same standard locality code as the photos and tissue specimens. Voucher specimens should be deposited in a museum at the conclusion or during a study paying attention to permit regulations regarding host country museums. Tissue specimens for DNA analysis should also be preserved in museums for future reference (i.e. use as little tissue as possible in your work so that future researchers can return to the same specimens). For Argentina you need collecting and export permits if you are depositing tissues in museums outside of the country (of course the primary depository is in the host country). It may be easier to arrange a formal loan of the specimen because the legal process for permanent deposits of tissues outside the country may take years.

N) If databasing is not done in the field, keep a running list at the back of your field notebook of all specimens collected. Collecting permits generally require a report soon after returning.

O) If possible database specimens in the field. If not, do this as soon as possible after returning. Note that if you have a laptop with you in the field, weather conditions are often available on line. Barcodes can be used to link specimen vials to information in a database. If you are doing collaborative work, consider using a common (cross-platform) database system such as Filemaker. Other databasing systems include: Biota,...

Back in the laboratory: develop an efficient scheme for keeping track of specimens in the freezer. Use controls and checks at every data collection step. Well-respected biologists have mixed-up the DNA's of well known organisms and even published bizarre conclusions and attempted to justify them. Some studies were later found to contain chimeric DNA—mixed sequences from two different organisms—or misaligned data. To the embarrassment of these researchers, peers caught the mistakes after publication. If little-known organisms were to be mixed up—especially closely related populations or species—the mistakes might never be discovered and incorrect conclusions would remain in the literature, possibly causing expensive problems for future researchers.

3) Data checks in the Extraction and PCR stage:

A) Aliquot various reaction components including extractions, primers, dNTP's, etc. Freeze/thaw cycles can cause degradation.
B) Label your tubes carefully.
C) Do every step in the same order.
D) Prepared blank PCR forms are useful for recording running conditions and amounts of the various components.
E) Always include positive controls (a specimen and primers you know to work) and negative controls (all PCR ingredients except DNA) with your PCR runs.
F) Know the size of piece to expect, i.e. know the locations of your primers (also know the degree of conservation of your primers so that you know whether to expect high probability of amplification.). Examine comparisons of primer sequences among species. Map your primers onto a diagram of the gene sequence; don't work blindly. Name all newly created

primers informatively (e.g. as in Simon et al. 1994) with gene name, location on a known sequence, and direction.
G) Assemble PCR components on a lab bench where no amplified DNA has ever existed and use separate pipettors for amplified DNASSD.
H) Dilute primers with special pipettors reserved for primer dilution only.
I) Pipette Taq with special pipettors reserved for Taq only.
J) Wear gloves.
K) IMPORTANT CONTROL: For species studies: extract more than one individual of each species and extract them on different days. Preferably, extract three or four individuals some from the same populations and others from the extremes of the range (cryptic species may exist). For deeper level studies, you may be able to substitute two species in the same genus for two specimens of the same species. For phylogeography studies, this is less of a problem because many more specimens are needed per population/species to estimate within and among population variation and achieve statistical significance.
L) The number of specimens needed for a phylogeography study may be more than can be easily collected or afforded. If so, during the first year of fieldwork, a wide sweep of populations can be made sampling only a few individuals per population. Genetic analysis of these specimens can provide a preliminary picture of variation that can guide future collecting and sequencing. Sample sizes in the middle of a genetic subdivision can be smaller than those at the edges that may contain contact zones and genotypes from other subdivisions.

4) In the sequencing stage:

A) Sequence in both directions. As a last resort, if you can't sequence in both directions for technical reasons, use overlapping segments primed in the same direction to check. This is a less desirable alternative because there may be errors specific to one direction.

B) For generic studies, sequence at least two individuals from each species and sequence them on different days.

C) Checking autosequencing results is an important step. The machines make predictable errors in calling that can be picked up by visual inspection. This is an essential step.

5) After Sequencing. Look at your sequence data! And align with care.

A) Calculate the nucleotide bias of your DNA and check to see if it is typical of your organism/organelle. If you suspect errors, re-extract, re-PCR, and re-sequence. (The cow that was mistaken for a wasp would have certainly been found if this simple check had been done.)
B) Make sure that all sequences read the same in both directions (often one primer works better than the other). Check discrepancies by re-sequencing.
C) Align sequences from multiple individuals from the same population or multiple populations of the same species to identify errors. In a protein-coding gene, all differences between sequences should be checked, especially second position mutations (all will change amino acids and this should be rare in closely related species).
D) Remember that bases on the ends of a sequence are more like to be wrong so don't be greedy when chopping off the ends.

E) Align RNA and intron data using information from conserved motifs and secondary structure.
F) Fold the RNA into the secondary structure to make sure that bases in helical regions pair properly.
G) Check conserved motifs to see if they are conserved (substitutions do occur in these regions but they are rare).
H) Align protein genes using amino acid translations (in all three frames if necessary).
I) Look for frame shifts or deletions that are not multiples of threes (these usually indicate errors). However, sometimes single-base frame shift mutations may be real but ignored during translation (Millbury and Gaffney 2005; Beckenbach et al. 2005).
J) Alignments should not be done by ordering taxa according to presumed phylogenetic relationship because this may introduce subjective bias. However, obvious taxa such as "plants" vs. "animals" or "fish" versus "birds" should be aligned in groups.

6) In the data analysis stage

For Phylogenetic Analysis: If your geographic structure is good, phylogenetic analysis of organelle genomes can be used for phylogeographic studies. Within well-defined sub-regions, phylogeographic techniques such as TCS parsimony and nested clade analysis can be used to study fine-scale population structure.

A) For phylogenetic trees: Check for nucleotide bias among taxa. None of the regularly used phylogenetic methods can accommodate this. This is unlikely to exist among closely related populations or members of a species complex.
B) If any biological/systematic information on your study organisms exists, use this to determine if any taxa are placed in strange positions on the tree. If so, check to see if any samples may have been mixed up. If a mix-up has occurred, sequence another specimen and compare its sequence to the problematic specimen. If no other specimen exists, at least try re-extracting and re-sequencing the original specimen.
C) Know the assumptions of each of your phylogenetic analysis methods. Compare the results of different models and different analyses to reveal potential biases in the data.
D) Postulate reasonable biogeographic (dispersal and vicariance) hypothesis that can be tested later using data from multiple taxa. This will allow the use of a priori (e.g., KH) rather than less powerful a posteriori (e.g., parametric bootstrap) tests of phylogenetic hypotheses (Goldman et al. 2000).
E) Parametric bootstrap tests can be used to test biogeographic hypotheses using multiple taxa (e.g., Carstens et al. 2005a).

References:

Ballard, J.W.O. and D.M. Rand. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Annu. Rev. Ecol. Evol. Syst. 36:621-642.

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

Carstens, B.C., S.J. Brunsfeld, J.R. Demboski, J.M. Good, and J. Sullivan. 2005a. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: Hypothesis testing within a comparative phylogeographic framework. Evolution 59 (8): in press.

Carstens, B.C., A. Bankhead III, P. Joyce, and J. Sullivan. 2005b. Testing population genetic structure using parametric bootstrapping and MIGRATE-N. Genetica 124:71-75.

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol.49:652-670.

Graham, C.H., S.R. Ron, J.C. Santos, C.J. Schneider, and C. Moritz. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in Dendrobatid frogs. Evolution 58: 1781-1793.)

Knowles, L.L. and Wayne P. Maddison. 2002. Statistical phylogeography, *Mol. Ecol.,* 11: 2623-2635.

Milbury CA, Gaffney PM. 2005. Complete Mitochondrial DNA Sequence of the Eastern Oyster, *Crassostrea virginica*. Marine Biotechnology 7:697-712.

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

Vellend, Mark. 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. Am. Nat. 166:199–215.

Vellend, M. and M.A. Gerber. 2005. Connections between species diversity and genetic diversity. Ecology Letters 8:767-781.

****If you have more ideas on data checking please contribute to our compilation***

Notes on permanent labels:

Date: Thu, 07 Feb 2002 17:53:22 +0000
To: bug-net@sfu.ca
From: Vince Smith <vsmithuk@yahoo.co.uk>
Subject: Replies re. insect DNA degradation and specimen labeling

Dear Bug-netters,

To summarize, most (but not all) agree that for semi permanent archival (up to 10ish years), laser printed labels will be okay. Printing on bleach free paper or treating it with some sodium
thiosulphate should help mitigate the effects of halogens in the paper damaging the DNA. However, for very long term storage, much more sophisticated systems printing on polyester

with a resin ink thermal printer are the best option (see Ted Schultz's e-mail). Most of the messages suggest that pencil labels are preferred and very permanent, but
we have too many samples and record too much data on our labels for this to be practical. I think for the moment we will go with laser printed labels printed on bleach free paper, backed up with an external sticky labels referencing just the specimen code.

Thanks again to all the who responded. I've placed all the messages on this subject at http://darwin.zoology.gla.ac.uk/~vsmith/protocols/labels.html, and if I receive any further mailing I'll append them to this list. In the mean time, messages I've received to date are pasted below.
Reply from Ted Schultz: I don't know about the paper or toner question but alcohol preserved material in the Smithsonian entomology collections are only labeled with certain inks because laser printer "inks" do not stay attached over long periods of time. I don't know details but you might want to check into permanence of your labels as well. Rick Wilkerson

I would seriously not use laser-printed labels, a because you might lose the ink, and b because they may bleach. I would use pencil labels, on bleach-free paper on the inside, and printed labels on the outside of your tubes. This may sound like a lot of work, but I have had one labeling disaster in my life and I won't have another. We label tens of thousands of tubes in this way! Also, think carefully about the quality of tubes you use. get good ones! Graham Stone

My personal experience has been that we label specimens with handwritten labels placed INSIDE the cryotube. The paper is archival linen stock, the inks are usually water/alcohol insoluble pigment-based inks. The preservative is 95% ethanol, and we store the tubes at -80°C or colder. We get good DNA out of these specimens. BUT I have always worried about this question. We do not put labels inside tubes containing DNA extracts or PCR reaction products, but then the only labels are numbers written on the outsides of the tubes -- this also bothers me. Solutions we're currently exploring: We are now printing EtOH vial labels on polyester stock with a resin ink ribbon on a "DataMax" thermal printer. This is supposed to be archival, permanent, and non-reactive. We are probably going to use the same printer to print sticky-backed barcode labels that will permanently go around the OUTSIDE of cryotubes containing DNA extracts. This number will reference records in a database containing all the label data. We are assured (by the supplier) that these sticky-backed labels are permanent even when immersed in liquid nitrogen. One more thing: Research here at our museum has shown (or so I am told) that laser-printed labels are NOT permanent or archival in EtOH. Printed on archival (acid-free) linen stock, they are fine for pinned or other dry-label use, but can be expected to degrade over time in alcohol. I am told that the toner will eventually come off, leaving a blank label! That's why we're moving over to the polyester film thing. It would be so much easier if we could generate our pin AND EtOH labels on the same printer, but, due to the above problem, we can't.

Notes on storing purified DNA EVOLDIR (To join EVOLDIR send a message to Brian Golding, Golding@McMaster.CA; messages are also posted on http://life.biology.mcmaster.ca/~brian/evoldir.html)

* I have kept DNA samples from Colias butterflies, either as resuspended from ethanol precipitation into demineralized water, or as lightly buffered Chelex extracts, in -70 to -80 C ultralow freezers for five years or more and had them amplify up in PCR as well as when they were first prepared. So my experience would suggest that DNA likes low temperature

just fine. What might have been problematic would have been repeated freeze- thaw cycles, which would cause a lot of shearing.

* For long-term storage the temperature of –20 freezers tends to fluctuate, causing ice crystals to tear DNA strands (especially long ones). Apparently this tearing does not happen with –80 or -160 degree freezers.

* The best way to keep your DNA for long-term storage is to precipitate it in ethanol... will be good forever. For working samples, 4 C is good enough for some time.