

Reconstructing the Phylogeny of Scolytinae and Close Allies: Major Obstacles and Prospects for a Solution

Bjarte H. Jordal¹

Abstract—To enable the resolution of deep phylogenetic divergence in Scolytinae and closely related weevils, several new molecular markers were screened for their phylogenetic potential. The nuclear protein encoding genes, CAD and Arginine Kinase, were particularly promising and will be added to future phylogenetic studies in combination with 28S, COI, and Elongation Factor 1 α . The combined analysis of multiple molecular markers and wide taxon sampling is expected to resolve many previously unresolved nodes in scolytine phylogeny, but a completely resolved topology seems dependent on the inclusion of a large number of morphological characters. A well founded phylogeny will provide a powerful framework for testing evolutionary hypotheses on habitat selection and reproductive biology.

Introduction

Bark and timber beetles in the weevil subfamily Scolytinae (Kuschel and others 2000) comprise nearly 6,000 species worldwide and constitute a significant factor in forest ecosystems. These insects receive increased attention, not only from forest entomologists concerned with forest health, but from evolutionary biologists fascinated by an unprecedented wide range of different biological and ecological features. Host plants and diets include most woody substrates imaginable, from living to dead trees and shrubs, dry to soggy substrates, large logs to herbs and tiny seeds, to the cultivation of asexual “ambrosia” fungi as food for adults and larvae. Their reproductive biology varies even more, including various forms of monogyny, harem polygamy, inbreeding by regular sibling mating, and parthenogenesis.

As a key to understanding how these features may have influenced diversification in Scolytinae, the generation of a baseline phylogenetic hypothesis is of paramount importance. Particularly interesting evolutionary transitions that can be tested in a phylogenetic framework include the multiple transitions between gymnosperm and angiosperm host plants and the evolution of ambrosia fungus symbiosis, mating systems (including the origin of various pheromones used in mate attraction), and haplodiploidy and paternal genome elimination in regularly inbreeding species. However, a phylogenetic reclassification of the Scolytinae and close relatives is still in its infancy without yet a clear picture of the relationship between most groups. Current classification (Wood 1986, Wood 1993) is based on rather few morphological characters not yet subject to a full cladistic analysis, although a smaller sample of scolytine taxa was recently analyzed by Kuschel and others (2000). They found a monophyletic Platypodinae nested inside Scolytinae, with Cossoninae

¹ Museum of Natural History and Archaeology, University of Science and Technology, Trondheim, Norway

as the closest outgroup. Wider taxon sampling and addition of molecular data will provide a more rigorous test for this hypothesis.

This project thus aims at closing many of the gaps in our current understanding of scolytine phylogeny and continues our recent phylogenetic work using DNA sequence data (Farrell and others 2001). Using such data to resolve deep divergence (for example, Cretaceous) is not straight forward; however, most genes demonstrate substitution rates not suitable for phylogenetic reconstruction at this level. Thus, one necessary step towards a prospective phylogeny of Scolytinae and allies includes screening for new molecular markers and sampling from a broad range of taxa. This paper reports preliminary data on evolutionary rates and phylogenetic properties of five little used genes in beetle phylogenetics and compares these data to previously used genes with respect to resolving deep divergences. The role for morphological data in conjunction with DNA sequence data is briefly discussed.

Methods

Several criteria were used to select genes for future sequencing and phylogenetic analyses of Scolytinae: 1) the amplification rate must be high, at least 70 percent of the samples must amplify to reduce the negative effects from large amounts of missing data in phylogenetic analyses; 2) substitution rates must be low, measured by sequence divergence and compared to the slowly evolving Elongation Factor 1 α (EF-1 α); and 3) phylogenetic signal must be high, judged by the capability of resolving likely sister relationship (based on 28S + EF-1 α or morphology).

Based on promising preliminary results from recent phylogenetic studies of bees, primer information for five different genes were downloaded from the web: Abdominal-A (Abd-a), RNA polymerase II (POL), Sodium-Potassium ATPase (NaK), CAD (rudimentary), and Arginine Kinase (ArgK) (see: <http://www.entomology.cornell.edu/BeePhylogeny>). PCR cocktails included Qiagen Hotstar taq and conditions were optimized on a MJ-200 gradient cyler with the following specifications: 95 °F for 15 min, then 40 cycles of 46 to 58 °F (60s), 72 °F (90s), final extension cycle (72 °F) of 10 min. Final optimal amplification temperatures were 50 to 52 °F (60s), except Abd-A, which did not amplify the correct gene products.

Results and Discussion

Screening for New Molecular Markers

PCR products of putative Abd-A, POL, NaK, CAD, and ArgK were sequenced and submitted to blast searches in Gen Bank. Correct gene products were confirmed in all cases except Abd-A. Characteristics of each verified gene fragment are listed below (see also table 1).

POL—the primers polfor2a and polrev2a amplified 822 base pairs from 43 percent of the samples (n=72). None of the sequences contained introns. Several sequences of Platypodinae had up to 33 amino acid substitutions compared to other Platypodinae and Scolytinae, suggesting paralogous copies. After excluding these sequences, the substitution rate was nevertheless surprisingly high (fig. 1) given that it is supposed to be one of the more conservatively evolving protein coding genes in insects (Danforth and others 2006). The high substitution rate was furthermore consistent through the full

Table 1—Characteristics of molecular markers used in phylogenetic studies of Scolytinae, including the new markers screened in this study.

Gene	Amp. rate ^a	Length ^b	Intron ^c	Signal nuc ^d	Signal aa ^e	Paralogs
12S	99	400	-	low	-	no
16S	98	500	-	low	-	no
COI	95	1200	0	low	high	yes ^f
18S	98	1900	-	low	-	no
28S	98	800	-	high	-	no
EF-1 α , C1	90	927	0-1	high	low	yes
Enolase, ni	20	687	0	low	moderate	yes
Histone H3	50	328	0	high	low	yes
NaK	50	713	1	low	low	no
POL	43	822	0	low	low	yes
CAD	73	900	1	moderate	high	no
ArgK	80	1120	1	high	moderate	no

^a Average proportion of samples that amplify.

^b The longest fragment sequenced of any Scolytinae (bp).

^c Uncorrected maximum divergence level in Scolytinae.

^d Phylogenetic signal from nucleotide sequences.

^e Phylogenetic signal from amino acid coded sequences.

^f Mitochondrial copies inserted into the nuclear genome (numts, pseudogenes).

length of the sequence, which impedes ready design of more specific primers. Taken together with the incongruent topology and limited support for clades

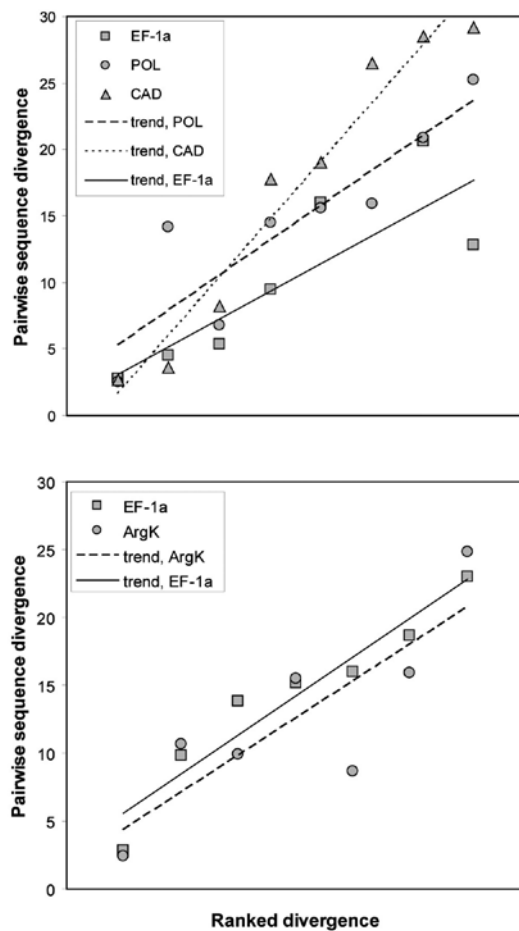


Figure 1. Sequence divergence (HKY corrected) between eight (above) or seven (below) species pairs of Scolytinae and Platypodinae, ranked by increasing divergence levels. Linear trend lines are added for illustrative purposes only.

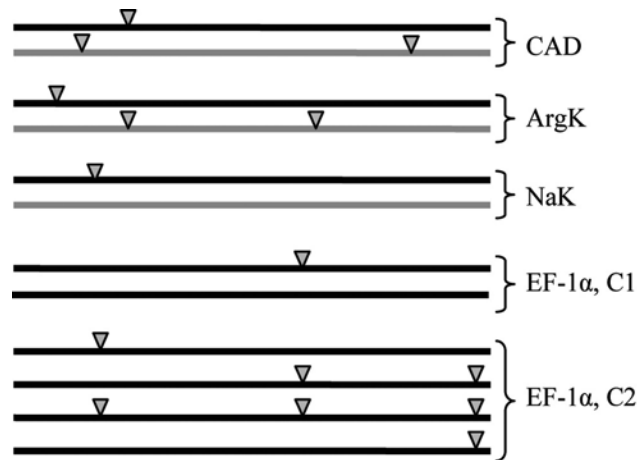


Figure 2. Intron maps for CAD, ArgK, NaK, and two copies of EF-1 α in Scolytinae. Gray lines indicate intron maps in bees (Halictidae). Triangles indicate the relative position of introns. Maps of different genes are not to scale.

that are otherwise strongly supported by EF-1 α and 28S data, this marker appears to have a limited phylogenetic potential in Scolytinae.

NaK—each of the primer pairs NaKfor1-NaKrev1, NaKfor1-NaKrev1a, and NaKfor2-NaKrev2 amplified multiple bands of various lengths in the majority of samples. The first primer pair amplified the correct product more consistently, amplifying 50 percent of the samples ($n=16$). The amplified fragment consisted of one intron close to the 5' annealing site (112 base pairs downstream from NaKfor1), ranging from 56 to 71 base pairs. There is no intron known from bees or from *Tribolium* (fig. 2). The substitution rate was high and ranged between 21 and 28 percent (p-distance), but a maximum of only five amino acid substitutions were found in the 642 base pair coding region. A phylogenetic analysis of six scolytine sequences was not at all consistent with previously well established clades. Taken together with the relatively low amplification rate and the high synonymous substitution rate, this gene seems less productive in higher level studies. It could, on the other hand, be an excellent marker for phylogenetic studies of genera and species complexes.

CAD—the primers 581F, apCADfor1, and apCADfor4, in conjunction with apCADrev1mod, produced band lengths of about 900, 700, and 455 base pairs. The correct product was amplified in 38, 56, and 73 percent of the screened samples ($n=40$), respectively, with very few multiple band patterns. One intron occurred in the two longest amplicons, close to the apCADfor1 annealing site. The intron ranged between 51 and 127 base pairs, but was universally missing in all Platypodinae in addition to one sequence of Scolytinae. Introns 5 and 6 in bees (Danforth and others 2006) were missing in these sequences. Several conserved primer sites provided ample opportunities to design more specific primers for consistent amplification of more than 800 base pairs (in progress). The substitution rate was higher than in POL and NaK (up to 32 percent, see also fig. 1), but even so the phylogenetic placement of scolytine and platypodine taxa was more consistent with the current classification. The Platypodinae was nevertheless slightly paraphyletic, but amino acid coding of these data resulted in a strongly supported, long branch of monophyletic Platypodinae. The amino acid data furthermore grouped distantly related genera in the tribes Corthylini and Polygraphini. The strong signal at the amino acid level may suggest that this gene holds a strong potential in separating tribes and subfamilies of various weevil groups. Dense sampling of genera and species within tribes will also probably increase phylogenetic accuracy for nucleotide analyses.

ArgK—the primers ArgKfor2 and ArgKLTfor3, in combination with the reverse primer ArgKLTrev2, produced band lengths of 1,120 and 490 base pairs, respectively. Very few samples amplified the longer fragment, while 80 percent of the samples amplified the shorter fragment (n=16). One intron was found close to the 5' end of the longest fragment, 93 base pairs downstream from the ArgKfor2 annealing site. Intron 1 and 2 in bees were not present in these sequences (fig. 2). The substitution rate was relatively low, not exceeding 22.2 percent for the taxa included (25.5 percent HKY corrected), which is about the same rate or lower than for EF-1 α (fig. 2). A phylogeny of 14 nucleotide sequences demonstrated high congruence with previously hypothesized clades (Ipini + [Dryocoetini + Xyleborini]; Platypodinae; *Cryphalus* + *Trypophloeus*; *Dendroctonus* + Hyalastini).

Building a Complementary Data Matrix

Several of the screened markers demonstrated sufficient amplification, evolutionary conservatism, and phylogenetic potential, to be included in the forthcoming higher level analysis of Scolytinae and close relatives (table 1). Particularly promising in this respect were CAD and ArgK. Each demonstrated consistent orthology and ample regions for primer design. These two gene regions will add some 1,500 to 1,700 base pairs to a matrix consisting of a mixture of conserved nucleotide sequences, rapidly evolving amino acid sequences, and morphological characters.

Among the numerous genes previously screened for scolytine phylogenetics, only a handful have proven useful in resolving deeper divergences (Farrell and others 2001, Sequeira and others 2000). Adding to the problem of rapidly evolving genes, the most extensively surveyed and slowly evolving gene, 18S ribosomal DNA, does not provide sufficient information to resolve relationships between most scolytines and other weevil groups (Marvaldi and others 2002). 28S seems more promising in this respect, with considerable contribution to combined data analyses (Sequeira and others 2000). A stronger phylogenetic signal is furthermore extractable from this gene if alignments are guided by secondary structure, potentially resolving the majority of scolytine tribes (Sequeira, pers. Com; Cognato and others, in progress).

Mitochondrial genes (12S, 16S, COI) are generally highly saturated by multiple substitutions in higher level studies. However, when COI nucleotides are translated into amino acids, there is apparently sufficient signal to resolve several important groups of Scolytinae (Sequeira and others 2001). Preliminary analyses of 223 COI amino acids (76 parsimony informative characters) from a wider range of taxa furthermore demonstrated a strongly supported Platypodinae nested within Scolytinae. Taken together with the benefits of DNA barcoding using COI nucleotides, the phylogenetic information in COI amino acids makes this gene a logical part of any large data set.

Among the nuclear protein encoding genes, *Enolase* has been used with some success (Farrell and others 2001, Jordal and Hewitt 2004). However, amplification of this gene is difficult, and the substitution rate is more suitable for generic level and below. *Histone* genes are more promising with respect to substitution rate, and the H3 copy is used frequently in metazoan phylogenetics. Scolytines nevertheless amplify paralogous copies, and the short length of the H3 copy (328 base pairs) increases the relative cost per base pair sequenced for this gene. The most promising marker for scolytine phylogenetics so far has been *Elongation Factor 1 α* , although not without problems. Sometimes, two copies of different length can co-amplify (fig. 1), and are

usually distinguished by different intron structures (Jordal 2002). However, adding EF-1 α sequences from a wider range of taxa raised some doubts about the integrity of intron structure in orthology assessment. Phylogenetic analyses nevertheless sort these copies into two distinct clades with intron variation, as shown in figure 2. These minor problems aside, EF-1 α alone can provide considerable signal and support for a wide range of tribes and genera (Farrell and others 2001, Jordal 2002, Sequeira and others 2001).

Based on the preliminary results presented here, and the brief review of markers used in scolytine phylogenetics, five gene fragments are selected for further phylogenetic analyses. The combination of EF-1 α , ArgK, CAD, 28S, and COI (amino acid) sequences will possibly enable resolution of currently unresolved nodes. However, support for some groups will most likely remain weak and the addition of morphological characters is needed to increase resolution and support. A data matrix consisting of 240 taxa and more than 150 morphological characters is currently in progress. Preliminary results based on 47 of these characters indicate a strongly supported Platypodinae, with *Schedlarius*, *Mecopelmus*, *Carphodicticus*, and *Coptonotus* as successive outgroups, and Scolytinae as the sister group to these taxa combined. These results are quite similar to Kuschel and others (2000) (see also Marvaldi and others 2002), but quite different from analyses of larval characters that relate Platypodinae to Dryophthorinae (Marvaldi 1997). The addition of molecular data and the remaining morphological characters will probably enable a definite test of these contrasting hypotheses.

Acknowledgments

The preliminary results presented here are part of a joint venture with A. Cognato, A. Sequeira, and B. Farrell. Larry Kirkendall collected some particularly valuable samples and reviewed the manuscript. Funding was provided by the Museum of Natural History and Archaeology, Trondheim.

References

- Danforth, B. N.; Fang, J.; Sipes, S. 2006. Analysis of family-level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II. *Molecular Phylogenetics and Evolution* 39: 358-372.
- Farrell, B. D.; Sequeira, A.; O'Meara, B.; Normark, B. B.; Chung, J.; Jordal, B. 2001. The evolution of agriculture in beetles (Curculionidae: Scolytinae and Platypodinae). *Evolution* 55: 2011-2027.
- Jordal, B. H. 2002. Elongation Factor 1 α resolves the monophyly of the haplodiploid ambrosia beetles *Xyleborini* (Coleoptera: Curculionidae). *Insect Molecular Biology* 11: 453-465.
- Jordal, B. H.; Hewitt, G. M. 2004. The origin and radiation of Macaronesian beetles breeding in Euphorbia: The relative importance of multiple data partitions and population sampling. *Systematic Biology* 53: 711-734.
- Kuschel, G.; Leschen, R. A. B.; Zimmerman, E. C. 2000. Platypodidae under scrutiny. *Invertebrate Taxonomy* 14: 771-805.
- Marvaldi, A. E. 1997. Higher level phylogeny of Curculionidae (Coleoptera: Curculionoidea) based mainly on larval characters, with special reference to broad nosed weevils. *Cladistics* 13: 285-312.
- Marvaldi, A. E.; Sequeira, A. S.; O'Brien, C. W.; Farrell, B. D. 2002. Molecular and morphological phylogenetics of weevils (Coleoptera, Curculionoidea): Do niche shifts accompany diversification? *Systematic Biology* 51: 761-785.

- Sequeira, A. S.; Normark, B. B.; Farrell, B. B. 2001. Evolutionary origins of Gondwanan interactions: How old are *Araucaria* beetle herbivores? *Biological Journal of the Linnean Society* 74: 459-474.
- Sequeira, A. S.; Normark, B. B.; Farrell, B. D. 2000. Evolutionary assembly of the conifer fauna: Distinguishing ancient from recent associations in bark beetles. *Proceedings of the Royal Society of London series B* 267: 2359-2366.
- Wood, S. L. 1986. A reclassification of the genera of Scolytidae (Coleoptera). *Great Basin Naturalist Memoirs* 10: 126.
- Wood, S. L. 1993. Revision of the genera of Platypodidae (Coleoptera). *Great Basin Naturalist* 53: 259-281.

current address:

Department of Biology, University of Bergen, Norway.