Increase Data Characters to Construct the Molecular Phylogeny of the *Drosophila auraria* Species Complex

Lu Gan1*, Gaodong Li1*, Wenhao Li1, Qingtao Zeng1, Yong Yang1,2#

1College of Life Science, Hubei University, Wuhan, China
2Hubei Collaborative Innovation Center for Green Transformation of Bio-Resources, Wuhan, China

Email: yangyong@hubu.edu.cn

Abstract

Previous phylogenetic analyses of the *auraria* species complex have led to conflicting hypotheses concerning their relationship; therefore the addition of new sequence data is necessary to discover the phylogeny of this species complex. Here we present new data derived from 22 genes to reconstruct the phylogeny of the *auraria* species complex. A variety of statistical tests, as well as maximum likelihood mapping analysis, were performed to estimate data quality, suggesting that all genes had a high degree of contribution to resolve the phylogeny. Individual locus was analyzed using maximum likelihood (ML), and the concatenated dataset (21,882 bp) were analyzed using partitioned maximum likelihood (ML) and Bayesian analyses. Separated analysis produced various phylogenetic relationships. Phylogenetic topologies from ML and Bayesian analysis based on concatenated dataset show that *D. subauraria* was well supported as the first species by separated analysis, concatenated dataset analysis, and some previous analysis, then followed by *D. auraria* and *D. biauraria*, *D. quadraria* and *D. triauraria*. The close relationships of *D. quadraria* and *D. triauraria* were consistent with most previous studies. The phylogenetic position of the *D. auraria* and *D. biauraria* will be resolved by more data sets.

Keywords

*Drosophila auraria* Species Complex, Phylogenetic Reconstruction, Multiple Genes

1. Introduction

The members of the *auraria* species complex in which ordinarily five members were involved (*D. auraria*, *D. biauraria*, *D. subauraria*, *D. quadraria* and *D. triauraria*)...
auraria) [1] [2] were considered as perfective model for reproductive isolation, flight activities, ability of diapaus, courtship songs and cold tolerance [3]. Recently, the phylogeny of the auraria species complex was studied based on various data, specifically, DNA sequence data. However, all analyses brought conflicting phylogenetic hypotheses [4]-[18] (Figure 1). The cause of the conflicting hypotheses is not known. All previous studies on the phylogeny of this species complex are based on different sample sizes or genetic markers. Differences in the number of taxa and the number of genes can have an effect on phylogenetic accuracy [19]. In many previous phylogenetic treatments of this species complex, representatives of only 4 species or less were included [4] [7] [13] [14] [15] [16] [17] [18]. Incomplete or insufficient taxon sampling has led to major inconsistencies in phylogenetic reconstructions [20] [21] [22] [23] [24]. On the other hand, differing sets of genetic markers were selected in previous studies, the most previous investigations were based on no more than two genetic markers [9] [11] [12] [15] [16] [17], phylogenetic hypotheses deduced from small amounts of sequence data would be incongruent or pool support [25]. Moreover, highly conserved genetic markers were involved in some analyses [3] [7] [13] [14] but some authors suggested that fast-evolving DNA regions were prior to analysis the molecular phylogenies of closely related species [26]. Although the phylogenetic relationships of these five members were deduced from 17 loci [18], the hypothesis that “increasing sampling outside the group may decrease accuracy” [27] may have applied; therefore, Yang (2012) did not resolve this complex problem. Many investigations suggested that maximizing gene numbers was advantageous to resolve complex phylogeny [28] [29]. Consequently, it

Figure 1. The diagram of the phylogenetic relationships of auraria species complex based on different data sets. (A, B, Q, S and T are D. auraria, D. biauraria, D. subauraria, D. quadraria and D. triauraria, respectively.
was advantageous to reconstruct the phylogeny of the five species based on increasing gene sampling sizes.

Finally, in this study, 22 genes segments were first used to reanalyze the phylogenetic relationships of D. auraria, D. biauraria, D. subauraria, D. quadraria and D. triauraria. These loci included partial genomic sequences of mitochondrial genes: cytochrome oxidase subunit I (COI), cytochrome oxidase subunit II (COII), mitochondrial genes ND1 (ND1) and ND4 (ND4); and nuclear ribosomal sequences: 28S rDNA (28S), internal transcribed spacer of nuclear ribosomal DNA (ITS including ITS1, 5.8S, 2S, and ITS2), and nuclear genes: amylase (amy), a parologue of the amylase genes (amr), sn-glycerol-3-phosphate dehydrogenase (gpdh), histone 2 spacers (h2s), Dopa decarboxylase (ddc), extra sexcombs (esc), hunchback (hb), extron 2, 3, 4 of alcohol dehydrogenase gene (adhl234), nucleoporin 96 - 98 gene (nup), membrane protein (patched) gene (ptc), and Xenopus Cdc6 (cdcl), genes for odorant-binding protein 57d, odorant-binding protein 57e (odo), multidrug-resistance associated protein 1-(mrp1), wingless (wgl), intron1 of bab gene (bab1), endophilin B (endoB).

2. Materials and Methods

2.1. The Study Taxa and Sequences Data

The sequences of COI, COII, ND1, ND4, 28S, ITS, amy, amryel, gpdh, h2s, ddc, esc, hb, adhl234, nup, ptc, Cdc6 were download from GneBank (GenBank accession numbers were listed in Yang et al., 2012). Sequences of odorant-binding protein 57d and e (odo), multidrug-resistance associated protein 1(mrp1), wingless (wgl), intron1 of bab gene (bab1) and endophilin B (endoB) were newly presented in this study. The detail information is given in Table 1. D. melanogaster was selected as the out group. PCR conditions and primers are listed in Table 2.

2.2. Sequence Alignment and Statistical Tests

Alignment of multiple DNA sequences was performed with MUSCLE for each gene [30]. The base composition, variable sites, and average genetic p-distance among all taxa were calculated by MEGA 4 [31]. The degree of nucleotide substitution saturation for each gene was tested using DAMBE 4.5.47 software [32]. A test for homogeneity of base frequencies across taxa was conducted using

<table>
<thead>
<tr>
<th>Species</th>
<th>GenBank Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>odo</strong></td>
<td><strong>mrp1</strong></td>
</tr>
<tr>
<td>D. auraria</td>
<td>EU835204</td>
</tr>
<tr>
<td>D. biauraria</td>
<td>AY465281</td>
</tr>
<tr>
<td>D. quadraria</td>
<td>AY465282</td>
</tr>
<tr>
<td>D. subauraria</td>
<td>AY465283</td>
</tr>
<tr>
<td>D. triauraria</td>
<td>AY465284</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>AE014297</td>
</tr>
</tbody>
</table>

Scientific Research Publishing
Table 2. PCR conditions and primers.

<table>
<thead>
<tr>
<th>Gene segment and length (bp)</th>
<th>Forward primers (5’-3’)</th>
<th>Reverse primers (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>odorant-binding protein 57d and e (~1070 bp)</td>
<td>OdF1: CTTTTGAATTACATTGCCGTA</td>
<td>OdR1: AATCCGGGTGCTTATAGC</td>
</tr>
<tr>
<td></td>
<td>OdF2: GTCTAAGCACGCGGAGT</td>
<td>OdR2: AGGATTGGAAGACGACGGAA</td>
</tr>
<tr>
<td></td>
<td>OdF3: TCCCGTGTCTTCAATCCCT</td>
<td>OdR3: CATCCAGATATTGAGGCA</td>
</tr>
<tr>
<td>multidrug-resistance associated protein 1 (~1070 bp)</td>
<td>MrpF1: TTATGCGGTTCCGAGCT</td>
<td>MrpR1: TTGGTCTGTCGCGGCATTCC</td>
</tr>
<tr>
<td></td>
<td>MrpF2: GGAATGCGCGCAAGACCAAA</td>
<td>MrpR2: AGCACAGAGGGTTCCACAGC</td>
</tr>
<tr>
<td></td>
<td>MrpF3: GCTGCCGCAAAAATTTGCT</td>
<td>MrpR3: TAGCGCGAGAGCCAA</td>
</tr>
<tr>
<td>wingless (~1070 bp)</td>
<td>WF1: GCTGGATGCGACTGGCAA</td>
<td>WR1: ACCTATTTATGTGGCGAACC</td>
</tr>
<tr>
<td></td>
<td>WF2: GGGCAGACCAAATATAGGT</td>
<td>WR2: GGCAGACGCTCCACACCAA</td>
</tr>
<tr>
<td>endophilin B (~1070 bp)</td>
<td>EbF1: GGAGGCCGGTTACCACGA</td>
<td>EbR1: AGGACTTGGCGACCGGC</td>
</tr>
<tr>
<td></td>
<td>EbF2: GCCGTGCGCAAGTTCCT</td>
<td>EbR2: ACTACAAGTAGGGCGGCA</td>
</tr>
<tr>
<td>intron1 of bab gene (~1070 bp)</td>
<td>BabF1: CACATAAAAAATCGCAACA</td>
<td>BabR1: TGCGGAGCAGCATGTGCAAC</td>
</tr>
</tbody>
</table>

PAUP 4.0 beta 10 [33].

2.3. Nucleotide Evolutionary Model Selection, Phylogenetic Analysis

For separate analysis, maximum likelihood (ML) trees for each locus were constructed in PAUP**v.4.0b10 [33] with the best nucleotide substitution model as determined by the Akaike Information Criterion (AIC). The concatenated dataset was divided into 22 partitions representing 25 genes, the best-known likelihood (BKL) tree for concatenated dataset was inferred after conducting 1000 RAxML runs using the f-d option for thorough searching and bootstrap replicates were performed in the multithread compiled version of RAxML-7.04. And Bayesian analysis running in MrBayes-3.1.2 [34] with 1,000,000 MCMC generations using the substitution model and parameters deduced from Model Test 3.06 [35].

2.4. Alternative Phylogenetic Hypotheses Test

SH test using CONSEL version 0.1 [36] were performed to test the statistical support of most of the previous hypotheses (Figure 1) and the hypotheses deduced from separate analysis. The BKL tree as optimal likelihood tree was modified using TreeView [37] to produce phylogenetic trees representing the alternative hypotheses.

3. Results

3.1. Sequence Alignment and Statistical Tests

Aligned sequences for the individual gene regions varied from 334 to 2455 bp in length, and the variation and parsimony informative sites were quite different among all genes, bab1 and 28S contain the highest and lowest number of parsimony informative sites, respectively (Table 3). Most of the average p-distances among the taxa were lower than 10% (18 out of 22); the mrp1 and bab1 have
larger values, whereas 28S, adh234, and odo have very small values (all lower than 2.2%). The test for substitution saturation [32] show that all gene regions have no substitution saturation. The sequences of all fragments show homogeneity of base frequencies (P ≥ 0.05).

3.2. Phylogenetic Analysis

The topologies of the trees deduced from the concatenated dataset under the ML and Bayesian analysis were completely identical (Figure 2), the five species in auraria species complex consisted of three lineages, the D. subauraria is the first species, then D. auraria and D. biauraria, D. quadraria and D. triauraria. The percent bootstrap support in ML analysis and posterior probabilities in Bayesian analysis all are 100 and 1.0, respectively. Maximum likelihood (ML) trees for each locus constructed in PAUP*v.4.0b10 [33] were different (Figure 3).

3.3. Alternative Hypotheses Test

All ML trees from each gene completely supported the melanogaster species group comprised of three monophyletic lineages: the ananassae subgroup, the montium subgroup, and the melanogaster subgroup plus oriental subgroups; however genes differed in the relationships among these groups. The montium subgroup was supported as the sister taxon of all remaining members of the melanogaster group by 6 of the 17 genes. The close relationships of the melanogaster, suzukii, and takahashii subgroups were supported by 4 genes. All 17 genes supported suzukii and takahashii as the sister lineages, and 7 genes supported the monophyly clade of ficusphila, eugracilis, elegans, and rhopaloa subgroups. Five of the 17 genes accepted the paraphyly of the suzukii subgroup, in which D. lucipennis is the sister species of D. elegans (see Table 3, and supplemental material). The p-values of all alternative phylogenetic hypotheses (Figure 1), except the hypotheses (p-values = 0.016) suggested by van der Linde and Houle (2008), are significantly lower than 0.005.

4. Discussion

Quality Evaluation for All the Representative Genes

Data quality is crucial for phylogenetic analysis, especially, based on DNA data-
Figure 2. The phylogenies of *auraria* complex were deduced from the concatenated dataset under the ML and Bayesian analysis. The number on the branch refer to bootstrap support in ML analysis and posterior probabilities in Bayesian analysis. *D. melanogaster* was out group.

Figure 3. Phylogenetic relationships of the *auraria* species complex were constructed based on different single locus.

sets; improper data which conceal conflicting evidence could lead to incorrect phylogenetic tree topologies [38]. All data in this study included various parsimony information sites, and the average genetic p-distance was close to 10% (Table 1), except for 28S (4.0%) and the four mtDNA (COI = 8.6%, COII = 8.1%, ND1 = 7.8%, and ND4 = 7.9%). 28S is 343 bp in length, a small part of the 28S complete sequence; the low genetic p-distance could come from arbitrary selection, hence, the conservative regions of this gene was selected. The mtDNA was traditionally considered as conservative genes; these kinds of genes were effective for discovering “higher-level” phylogenies. The likelihood mapping (Table 3) results also indicated that these four genes have significant contributions to resolve the phylogeny of the *melanogaster* species group (COI = 80.0%, COII =
81.5%, ND1 = 88.8%, and ND4 = 89.9%).

According to the test method described by Xia et al. (2003), most of the genes showed no saturation (Table 2) except for ITS and nup, with little saturation. Although saturation could bring noise into the phylogenetic analysis, these two genes included useful phylogenetic information, and the values of the average genetic p-distance suggest that these two genes have “fast” evolutionary rates. Thollesson (1999), however, suggested that fast genes also included phylgetic signal and Scholler (1994) demonstrated divergent genes, e.g., ITS, are appropriate genetic markers to discover the phylogeny of the melanogaster species group. The likelihood mapping also showed these two genes have considerable amounts of phylogenetic signal. Therefore, these genes were included to avoid loss of phylogenetic signal.

The concatenated dataset has a large amount of useful parsimony information (parsimony sites, 4365), and the average genetic p-distance (14.0%), together with the likelihood mapping values (resolved quarters, 99.5%), indicated that the combined data has provided sufficient phylogenetic signals to resolve the phylogenetic relationships of the melanogaster species group.

In the present study, the phylogenetic relationships of the auraria species complex were reconstructed based on 22 gene segments (Figure 2). The phylogenetic tree indicated that D. subauraria is the first species, then D. auraria and D. biauraria, D. quadraria and D. triauraria, which were supported with high bootstrap values and posterior probabilities (100 and 1.0, respectively). D. subauraria was absent in the phylogenetic analysis before it was found in Japan [39], but based on the RAPD data, it was the first species in this species complex. The phylogenetic position of the D. subauraria as the first species was also supported by some previous analysis [3] [5] [8] [10] and half of the separated analysis (11 genes of the 22 genes. On the other hand, no evidences show that D. subauraria could produce fertility offspring with other member of this species complex. All the above mentioned evidences suggested that the D. subauraria was the earlier divergent species. D. quadraria was always considered as the first species before D. subauraria was taken into phylogenetic analysis [8], however, D. quadraria and D. triauraria could produce cross-fertilize generations [2], which indicated that there was no reproductive isolation between these two species, therefore, these two species were ever treated as the same species [1] [2]. Moreover, some previous analysis based on two-dimensional electrophores and DNA data sets [9] [15] supported the close relationships of D. quadraria and D. triauraria, which were also supported by sixes gene data sets in the analysis (abb1, edo, ND1, ptc, amy, and COII). The phylogenetic relationships of D. auraria and D. biauraria were only supported by three gene data sets in separated analysis (ptc, abb1, and mrp1) and some previous evidences [4] [18]. From all the analysis, it could be draw a primary conclusion that D. subauraria was the first species, and D. quadraria and D. triauraria were the close relative two species, but the phylogenetic positions of D. auraria and D. biauraria should be resolved based on more data sets. However, in the future analysis of the phylogenetic relationships
of these species, it is advantageous to consider add more DNA sequences data. From all the separated and most previous analysis, it was obvious that any signal gene data was limited to discover the phylogenetic relationships of this species complex because almost all the previous analysis and separated analysis were rejected by analysis based on the concatenated dataset. Especially, some authors emphasized that phylogenetic hypotheses.

5. Conclusions

1) The phylogenetic relationship of Auraria species complex were analyzed based on different data, *D. subauraria* is the first species, *D. auraria* and *D. biauraria*, *D. quadraaria* and *D. triauraria* as the second clusters.

2) To discuss the phylogenetic relationship of relative species, it was advantageous to concatenate large dataset than single locus.

Acknowledgements

This work was supported by the Department science of Hubei Province (2012FFB00304) and the Foundation of Wuhan City Technology Bureau (2013070104010013). We thank the Drosophila Genetic Resource Center and the Bloomington Drosophila center for supplying the flies used in this study.

References


[3] Miyake, H. and Watada, M. (2007) Molecular phylogeny of the *Drosophila auraria* Species Complex and Allied Species of Japan Based on Nuclear and Mitochondrial DNA Sequences. *Genes & Genetic Systems*, 82, 77-88. [https://doi.org/10.1266/ggs.82.77](https://doi.org/10.1266/ggs.82.77)


Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/
Or contact ojgen@scirp.org