Molecular Phylogenetics and Evolution 57 (2010) 678–686

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Domestication of olive fly through a multi-regional host shift to cultivated olives: Comparative dating using complete mitochondrial genomes

F. Nardi [a,*], A. Carapelli [a], J.L. Boore [b], G.K. Roderick [c], R. Dallai [a], F. Frati [a]

[a] Department of Evolutionary Biology, University of Siena, Via Aldo Moro 2, 53100 Siena, Italy
[b] Genome Project Solutions, Inc., 1024 Promenade Street, Hercules, CA 94547, USA
[c] Department of Environmental Sciences Policy and Management, 137 Mulford Hall, University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 12 February 2010
Revised 9 August 2010
Accepted 10 August 2010
Available online 17 August 2010

Keywords:
Olive fly
Olea
Host shift
Mitochondrial genome
Dating
Coalescence
Diptera
Phylogeny
Agricultural pest

ABSTRACT

The evolutionary history of the olive fly, *Bactrocera oleae,* was reconstructed in a phylogenetic and coalescent framework using full mitochondrial genome data from 21 individuals covering the entire worldwide distribution of the species. Special attention was given to reconstructing the timing of the processes under study. The early subdivision of the olive fly reflects the Quaternary differentiation between *Olea europea* subsp. *europea* in the Mediterranean area and the two lineages of *Olea europea* subsp. *cuspidata* in Africa and Asia, pointing to an early and close association between the olive fly and its host. The geographic structure and timing of olive fly differentiation in the Mediterranean indicates a clear connection with the post-glacial recolonization of wild olives in the area, and is irreconcilable with the early historical process of domestication and spread of the cultivated olive from its Levantine origin. Therefore, we suggest an early co-history of the olive fly with its wild host during the Quaternary and post-glacial periods and a multi-regional shift of olive flies to cultivated olives as these cultivars gradually replaced wild olives in historical times.

1. Introduction

The olive fly, *Bactrocera oleae* Rossi, is a species of great economic importance throughout the world where olives are grown (Daane and Johnson, 2010). Its larvae are oligophagous and feed upon the pulp of fruits of the genus *Olea*, both wild and domesticated forms (Tzanakakis, 2006), causing significant losses to olive and oil production. The species has a long record of range expansions and invasions which have been well studied from perspectives of both population biology and phylogeography (Ochando and Reyes, 2000; Ochando et al., 2003; Augustinos et al., 2002, 2005; Nardi et al., 2003, 2005, 2006; Segura et al., 2008; Zygouridis et al., 2009). The olive fly is generally regarded as a Mediterranean species, being associated with one of the most historically and economically important crops of the area – the olive tree. However, neither olive fly nor olive originated in the Mediterranean region. The origin and ancient history of the olive tree is still matter of debate, but its basal diversification occurred most likely following the aridification of African midlands at the beginning of the Pliocene, with the separation of African and Asian lineages of *Olea europea* subsp. *cuspidata* and the North African/European *O. e.* group of subspecies. The dominant and pristine form of olive tree in the Mediterranean predating olive domestication was *O. e.* subsp. *europea* var. *sylvestris*, or wild olive (Besnard et al., 2007b, 2009; Baldoni et al., 2002). Pre-Quaternary in origin (Zohary and Spiegel-Roy, 1975; Besnard et al., 2007b) and once widely distributed, wild olives are now rare being displaced by cultivated forms (Lumaret et al., 2004). Despite their reduced pulp and limited oil content, wild olives were exploited by man during the Neolithic (Zohary and Hopf, 2000) and constitute suitable hosts for the olive fly (Katsoyannos, 1992; Tzanakakis, 2006). Domestication of such wild olive forms took place in the Levant around the 4th millennium BC (Zohary and Hopf, 2000; Lumaret et al., 2004) giving rise to the olive tree as currently known, *O. e.* subsp. *europea* var. *europea*, or cultivated olive. This variety was subsequently introduced through human-mediated commercial exchanges in most Mediterranean countries (Ruiz Castro, 1948; Boardman, 1976) to reach its present distribution and population density in the Middle Ages. As a result, two distinct olive populations have characterized the Mediterranean, a first of wild olives in the Quaternary and a second of domesticated olives in historical times.

The olive fly is found today in East and South Africa, the Mediterranean area, and Pakistan, largely matching the natural geographic range of wild olives up to a certain longitude (Tzanakakis, 2006).

* Corresponding author. Fax: +39 0577 234476.
E-mail addresses: nardifra@unisi.it (F. Nardi), carapelli@unisi.it (A. Carapelli), JLBoore@GenomeProjectSolutions.com (J.L. Boore), roderick@berkeley.edu (G.K. Roderick), dallai@unisi.it (R. Dallai), frati@unisi.it (F. Frati).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.08.008

In recent times, the fly has been accidentally introduced to California (Rice, 1999; Rice et al., 2003; Taylor, 2000; Nardi et al., 2005, 2006; Zygouridis et al., 2009).

In spite of the fly's worldwide distribution, the monophyly of the species has never been questioned, and three genetic groups have been described that correspond to the three aforementioned regions (Nardi et al., 2005). The Pakistani group, originally described as *B. oleae* var. *asiatica* (Silvestri, 1916) most likely represents a relict isolated population (Nardi et al., 2005). The more typical form of olive fly, found in East to North Africa and in the Mediterranean, seems to have evolved through a series of events of range expansion from Africa to the Mediterranean area (Nardi et al., 2005) and possibly westward in the latter region, as suggested by the observed gradual decrease in heterozygosity in an E–W cline (Augustinos et al., 2005). The current Mediterranean population is large and characterized by extensive gene flow, with some traces of differentiation arising only on a large geographic scale along an E–W axis and possibly, based on a single account of a Tunisian sample, across the Northern and Southern shores of the Mediterranean basin (Augustinos et al., 2005; Nardi et al., 2005; Segura et al., 2008; Zygouridis et al., 2009). Based on phylogeographic interpretation, a Quaternary origin was supposed for the Africa/Mediterranean diversification of olive fly (Besnard et al., 2007b), while the more recent colonization of the Mediterranean area was assumed to be subsequent (Augustinos et al., 2005; Segura et al., 2008; Zygouridis et al., 2009) or concurrent (Ruiz Castro, 1948; Nardi et al., 2005) to olive domestication and expansion in the area. The alternative possibility of the olive fly being present on wild olive forms predating the spread of domesticated olives, though equally viable, has never been considered seriously, apart from an interesting hypothesis of a secondary specialization of the olive fly on olive (Goulielmos et al., 2003).

While the general pattern of the olive fly's current genetic structure and routes of range expansions are reasonably well understood, the temporal aspects of these processes have been difficult to uncover due to the lack of suitable genetic markers. In this study, we examine directly this further dimension of olive fly evolution by assessing whether the time scale of olive fly expansion to and across the Mediterranean is compatible with an old colonization on wild olives, pointing to a passive dispersal and differentiation, or alternatively with a recent colonization associated with human-mediated olive tree domestication and diffusion. Accordingly, we explicitly test the hypothesis of a single event of a host shift from wild to cultivated olives, possibly connected to olive domestication, versus a multi-regional shift of olive fly onto cultivated olives (Supplemental Fig. 1).

Mitochondrial sequences have been exploited extensively to study evolutionary processes at the species level (Avise, 2000; Zink and Barrowclough, 2008 for a recent debate). Previous work showed that while single gene sequences could resolve major subdivisions of the olive fly worldwide (Nardi et al., 2003, 2005), these did not provide enough variability to allow for estimation of branch lengths and the temporal dimension of underlying processes, suggesting the potential utility of complete mitochondrial genomes. Because of the higher number of mutations observed, complete mitochondrial genomes allow for the use of more complex and time-aware methods of analysis (Carr and Marshall, 2008) and are arguably the best choice at present to maximize information content in terms of time estimates. Alternative approaches, such as sequencing and dating of multiple nuclear intron loci, have been difficult to implement thus far in this species. Recognized drawbacks of the mitochondrial genome are that it is inherited as a single locus, thus not allowing for comparative analyses nor for the intrinsic random nature of coalescence to be taken into account explicitly, that it may be subject to a certain degree of selection (Ballard and Whitlock, 2004) and that the stability of rate estimates might not hold for very recent nodes (<200 Ky to 1 My), an issue recently gaining attention as "time dependency of molecular rates" (Ho et al., 2005; Burridge et al., 2008; Papadopoulou et al., 2010; see a counter criticism by Emerson (2007)).

In order to address these limitations, we did not analyze the genome tree in terms of possible population groups, but rather used the tree and associated datings to propose a time scale for major and well supported population groups described previously using multilocus nuclear microsatellite data (Augustinos et al., 2005; Nardi et al., 2005; Zygouridis et al., 2009). Furthermore, only dates >500 Ky are discussed, and comparisons are made across time scales that differ by 1–2 orders of magnitude, thus limiting the possible effects of biases in time estimates over the biological interpretation of results.

2. Materials and methods

2.1. Diptera tree

Due to the lack of suitable calibration points for the olive fly tree, a dated tree of Diptera was produced using paleontological calibration points and the dates obtained for relevant nodes were subsequently used to calibrate the olive fly tree.

All complete mitochondrial genomes available at the time of writing from Diptera were associated with the two newly obtained sequences of *Bactrocera tryoni* and *B. oleae* specimen KEN_1008 (Supplemental Table 1). PCGs were retro-aligned using BioEdit (ver. 7.0.0: Hall, 1999), rRNAs were aligned using ClustalX (ver. 2.0.9: Larkin et al., 2007) and manually corrected, tRNAs were aligned based on secondary structure information. All 37 single gene alignments were concatenated to give a final dataset of 25 sequences by 14974 aligned positions. Sequences of Nematocera were used for outgroup comparison in all analyses.

Optimal partitioning strategy was assessed using a comparison of Bayes factors. MrBayes (v.3.1.2: Ronquist and Huelsenbeck, 2003) paired runs of four chains were conducted for four million generations sampling trees every 1000 generations, with burnin assessed by monitoring parameter traces and standard deviation of split frequencies across duplicate runs. Nine partitioning schemes were tested, composing partitioning by codon (1st/2nd/3rd, 1st + 2nd/3rd, 1st + 2nd + 3rd[1]; RNAs added as a further partition) and by strand (J/N, J/N with all parameter linked except for base frequencies, J + N). The best model of sequence evolution for each partition was selected according to MrModeltest (v.2.3: program available from the author at http://www.abc.se/~nylander/) and AIC criterion. Considering that one outgroup sequence (*C. arakawae*) always displayed strong rate acceleration (data not shown), this sequence was removed from all subsequent analyses.

A Bayesian dating method implemented in BEAST (v.1.4.8: Drummond and Rambaut, 2007) was used to obtain a dated Diptera tree based on optimal partitioning and models of evolution (1st + 2nd/3rd and J/N with all parameters unlinked; GTR + I + Γ model for all partitions except HKY + I + Γ for third positions of PCGs encoded on the J strand). In all analyses, strong priors were set upon the age of the three basal nodes corresponding to Brachycera, Cyclorrapha and Schyzophora according to paleontological and molecular data (Table 1; Wiegmann et al., 2003; Mostowski, 2000; Krzeminski and Evhenuis, 2000; Grimaldi and Cumming, 1999; McAlpine, 1970). A branching prior was set under a Yule process model and a relaxed molecular clock was assumed using

[1] *Abbreviations used:* My, millions of years; Ky, thousands of years; HPD, high probability distribution; s/s/My, substitutions per site per million years; 1st/2nd/3rd, first, second and third codon positions in protein-coding genes; J/N, strand of the mitochondrial genome where the majority/minority of genes is encoded, equals *plus*/*minus* strand; COX1, gene encoding for mitochondrial cytochrome *c* oxidase subunit 1.