EVOLUTION & DEVELOPMENT 12:3, 288–295 (2010)
DOI: 10.1111/j.1525-142X.2010.00414.x

Catching the phylogenic history through the ontogenic hourglass: a phylogenomic analysis of *Drosophila* body segmentation genes

Amir Yassin*, E. Kurt Lienau, Apurva Narechania, and Rob DeSalle
Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th St. at Central Park West, NY 10024, USA
*Author for correspondence (email: ayassin@amnh.org)

SUMMARY The phylogenetic information content of different developmental stages is a long-standing issue in the study of development and evolution. We performed phylogenetic analyses of 51 body segmentation genes in 12 species of *Drosophila* in order to investigate the impact of the mode of evolution of development on phylogeny inference. Previous studies of these genes in *Drosophila* using pairwise phenetic comparisons at the species group level revealed the presence of an "hourglass model" (HG), wherein mid-embryonic stages are the most evolutionarily constrained. We utilized two character-based approaches: taxonomic congruence using the relative consensus fork index (RCFI), in which phylogenies are inferred from each gene separately and compared with a total evidence tree (TET), and partitioned simultaneous analysis using several indices such as branch support (BS) and localized incongruence length difference (LILD) test. We also proposed a new index, the recapitulatory index (R), which divides the number of synapomorphies on the total number of informative characters in a data set. Polynomial adjustment of both BS and R indices showed strong support for the hourglass model regardless of the taxonomic level (species subgroup vs. subgenera), showing less phylogenetic information content for mid-developmental stages (mainly the zygotic segment polarity stage). Significant LILD scores were randomly distributed among developmental stages revealing the absence of differential selective constraints, but were significantly related to chromosomal location showing physical (linkage) impact on phylogenetic incongruence. RCFI was the most sensitive measure to taxonomic level, having a convex parabola at the species subgroup level in support of the hourglass model and a concave parabola at the subgeneric level in support of the adaptive penetrance model. This time-dependent discrepancy of best fit developmental model parallels previous conflicting results from the vertebrates. Because of the quasi-phenetic nature of this index, we argue that the discrepancy is due to the evolutionary rate heterogeneity of developmental genes rather than to fundamental differences among organisms. We suggest that simultaneous character-based analyses give better macroevolutionary support to the hourglass model of the developmental constraints on genome evolution than pairwise phenetic comparisons.

INTRODUCTION

The strong similarity between development (ontogeny) and evolution (phylogeny) has fascinated biologists since the dawn of the evolutionary theory (Gould 1977). Both processes imply the generation of advanced (derived) characters from primitive (ancestral) ones, though they differ in their generational time scale. The third law of von Baer (1828) stated that embryonic characters of different species progressively diverge from one another during ontogeny. Haeckel (1866) placed this law into an evolutionary context by formulating the biogenetic law according to which "ontogeny recapitulates phylogeny." A direct systematic application of this law was the preferential use of characters of early embryonic stages (thought to reflect ancestral forms) for inferring phylogenies than of those of late and adult stages. For Haeckel (1866), embryology was a tool to establish natural classification, a science he gave the name of *Systematische Phylogenie*.

The integration of comparative embryology within a phylogenetic context has given rise to the new discipline of evo-devo, in which the reconstruction of ancestral embryonic stages is crucial for the understanding of the evolution of development. For Haeckel, embryology was a tool to infer phylogenies, while in evo-devo phylogenies are the conceptual framework to understand the evolution of development. One of the major findings of evo-devo was that different embryonic stages do not diverge equally among species. Indeed, early and late stages are more divergent than middle ones that have been dubbed the "phylotypic stage" (Sander 1983). This has been attributed to fundamental epigenetic consequences

 © 2010 Wiley Periodicals, Inc.

of mid-embryonic stages on adult bauplan, resulting in an "hourglass model" of ontogenetic divergence with a constriction of phenotypic and genetic diversity at middle stages (Raff 1996). However, the generality of this developmental model is still questionable.

Recently, many studies have addressed this question at the molecular level, that is, studying sequence divergence of genes expressed at different developmental stages (Davis et al. 2005; Hazkani-Covo et al. 2005; Hanada et al. 2007; Cruickshank and Wade 2008; Roux and Robinson-Rechavi 2008). Such an approach is of course a convenient proxy because functional developmental divergence can also be related to epigenetic factors independent of the gene sequences (e.g., regulation of expression of developmental genes and transcriptomic modifications), whereas the latter are likely to affect developmental divergence. Within this conceptual frame, the ontogenic hourglass was well illustrated in the degree of conservation of body segmentation genes in Arthropods (Peel et al. 2005). In *Drosophila melanogaster*, these genes can be classified under five successive classes: maternal, gap, pair-rule, segment polarity, and segment identity (Hox) genes. The latter four classes are zygotically expressed. Comparison of these genes between *D. melanogaster* and its closely related species *D. simulans* from which it has diverged circa 2–4 Ma (Russo et al. 1995), revealed a general pattern at the intra- and interspecific levels: diversity is higher in early maternal genes than in later zygotic classes of which segment polarity genes show the lowest diversity (Cruickshank and Wade 2008). Davis et al. (2005) have previously shown that this hourglass pattern also persists when the same set of genes were compared between *D. melanogaster* and a more distant species, *D. pseudoobscura*, from which it has diverged circa 20–40 Ma (Russo et al. 1995; Tamura et al. 2004).

The present work aims to investigate the impact of the hourglass model on phylogeny inference, that is, how much are different embryonic stages phylogenetically congruent? We use segmentation genes of the 12 *Drosophila* species for which whole genome sequences have recently been published (*Drosophila* 12 Genomes Consortium 2007). The phylogeny of these species is well-established from different sources of evidence including morphology and complete sequences of nuclear and mitochondrial genomes (*Drosophila* 12 Genomes Consortium 2007; O'Grady and DeSalle 2008), allowing ad hoc homology assessment.

MATERIALS AND METHODS

Data matrix preparation and partitioning

Fifty-one genes involved in the development of body segmentation in *Drosophila* were sampled and were classified into five successive developmental stages: maternal (17 genes), gap (13 genes), pair rule (8 genes), segment polarity (5 genes), and segment identity (8 genes), according to the 55th edition of the Inter*active* Fly database (Brody 1999) visited on April 2009. For each gene, orthologous amino acid sequences of the 12 *Drosophila* species and chromosomal location in the *D. melanogaster* genome were obtained from the FlyBase database (Tweedie et al. 2009). Sequences for each gene were individually aligned using ClustalW (Thompson et al. 1994) with default parameters as implemented in MEGA version 4 (Kumar et al. 2008). Alignments were then concatenated into a larger matrix with three levels of partitions: first, for each individual gene; second, for each developmental stage; third, for each pattern of expression (maternal vs. zygotic); and finally, the whole data set without partitions.

Phylogenetic analysis

Maximum parsimony (MP) was used in phylogenetic analyses because of the few number of taxa (12 species) and suitability for the analysis of partitioned data. Phylogenetic analyses were conducted separately on each of the three levels of partitions generating 58 phylogenetic trees using close neighbor interchange algorithm with random addition of 10 sequences as implemented in MEGA. The tree inferred from the unpartitioned matrix containing the whole data set is the total evidence tree (TET), which is considered the reference tree (Fig. 1). TET was in concordance with the known phylogeny of the 12 species inferred from the complete sequence of their genomes (*Drosophila* 12 Genomes Consortium 2007).

Taxonomic congruence analysis

The major aim of this study is to test whether genes expressed at different developmental stages differ in their levels of phylogenetic corroboration. Two phylogenetic approaches are known in measuring corroboration: taxonomic congruence and total evidence. In taxonomic congruence, corroboration among data sets for a particular node is indicated by replication of that node in topologies derived from the separate data sets (Miyamoto and Fitch 1995). We compared phylogenies separately inferred for each individual gene with TET and estimated the relative consensus fork index (RCFI) which gives the proportion of nodes in common between the gene tree and the TET (Colless 1980). Hence, an RCFI of 0 indicates no single agreement between the gene tree and the species (TET) tree; whereas a value of 1 means that the gene tree faithfully replicates TET. Stability of a certain node (clade) can be estimated by counting the number of times it has been recovered from the separate analyses.

Phylogenetic incongruence analysis

In the total evidence approach, a comparison of individual-gene topologies is not required. Different measurements of phylogenetic incongruence have been developed; the most common ones are branch support (BS; Bremer 1988) and incongruence length difference (ILD; Farris et al. 1995). Each has originally been developed to estimate net incongruence (i.e., character conflict) among data sets in simultaneous analysis. A number of derivatives for each measurement have been proposed to estimate the relative support provided by each data set at a particular node. Both measurements and their derivatives were estimated here for the three levels of data partitions.

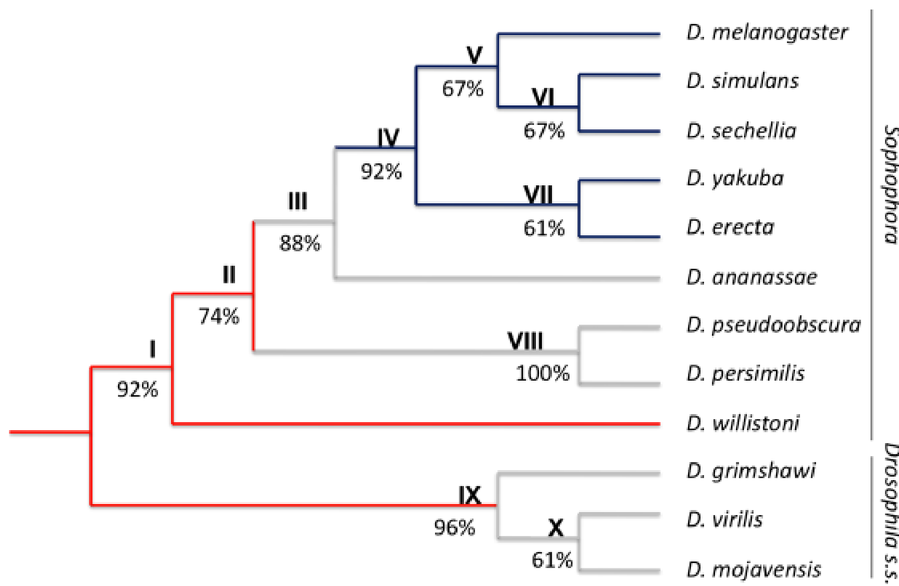

Fig. 1. Total evidence phylogenetic tree of 12 *Drosophila* species inferred from the simultaneous analysis of 51 developmental genes. Nodes are numerated using Latin numbers. Percents below nodes refer to the relative recovery of the node when the phylogeny was inferred individually from each gene. Branches of the two taxonomic levels used in the analyses are colored: the *melanogaster* species subgroup level (in blue) and the subgeneric level (in red).

BS is the minimum number of character steps for a data set on the shortest topologies that do not contain a particular node, minus the minimum number of character steps for that data set on the shortest topologies that do contain that node. BS can be positive if the node is supported by the data set, or zero or negative if the node is not supported. Baker and DeSalle (1997) proposed the partitioned branch support (PBS) index, which is an extension of BS, in which for any node the sum of PBS scores for the different partitions equals BS at that node. For a particular combined data set, a particular node, and a particular data partition, PBS is the minimum number of character steps for that partition on the shortest topologies for the combined data set that do not contain that node, minus the minimum number of character steps for that partition on the shortest topologies for the combined data set that do contain that node. Gatesy et al. (1999) noted that the interaction of different partitions in simultaneous analysis often implies hidden character support and conflicts. They developed the hidden partitioned branch support (HPBS) index, which is the difference between PBS and BS at a node for a partition.

ILD is a test detecting significant incongruence in the phylogenetic trees derived from different data partitions. It is composed of two parts, a test statistic (i.e., the additional homoplasy that arises when combining different partitions in a simultaneous analysis) and a null distribution generated by a randomization model. Such as BS, ILD is a summary measure for the entire tree. Thornton and DeSalle (2000) derived a metric called localized incongruence length difference (LILD) to quantify the conflict between partitions for any node in a phylogeny. LILD measures the tree length difference between a tree search, wherein a node optimal in TET constrains tree topology and a tree search where that constrain is lifted. LILD was calculated for each of the partitions at every node in TET. For the 10 nodes in a phylogeny of 57 partitions, this translates into 570 unique experiments. The nonparametric statistical significance of these LILDs was determined by evaluating the difference between them and a distribution of 1000 random partitions created from the concatenated TET matrix using the Bioperl script described by Narechania et al. (2005).

Homology content and the recapitulatory index

Variable characters in a matrix can be synapomorphies (characters derived in a common ancestor), autapomorphies (characters derived in a single taxon) or homoplasies (characters derived independently). Only synapomorphies (S) are indicators of the true phylogenetic history, whereas homoplasies (H) convolute the inference of this history. Autapomorphies are not relevant in phylogeny reconstruction using MP. The ratio of the number synapomorphies to the total number of phylogenetic informative characters (i.e., synapomorphies+homoplasies) can be considered a good measurement of the phylogenetic signal. We called this ratio the recapitulatory index $R = S/(S+H)$, estimated it for each individual gene and averaged for every developmental class. Because it is a ratio (and not a direct estimate of S), it is not expected to be affected by either the number of changes or the length of the sequences.

Homology was thus assessed for each of the 45,248 amino acid residuals used in the concatenated matrix in light of TET. The consistency index (CI) which is the minimum amount of change in the character divided by the length (number of steps) required by the character on the TET (Kluge and Farris 1969) was estimated for each residue. Nonetheless, CI cannot distinguish between synapomorphies (characters derived in a particular ancestor) and autapomorphies (characters derived in a single taxon) as both will have a CI value of 1. Hence, the retention index (RI), which is the difference between the maximum possible amount of change and the length required by the character divided by the difference between the maximum possible and minimum amounts of change (Farris 1989) was used. A synapomorphic character will have a RI value of 1, whereas RI for an autapomorphic residual is undefined. Both CI and RI were estimated using the PAUP* software package version 4 (Swofford 2003).

Model testing and taxonomical hierarchy

Poe and Wake (2004) recognized four evolutionary models of development: (1) the early conservation model (EC), wherein late developmental stages are less constrained than earlier stages (von Baer 1828); (2) the hourglass model (HG) wherein mid-developmental stages are more constrained than early and late stages (Raff 1996); (3) the adaptive penetrance model (AP) wherein mid-developmental stages are less constrained than early and late stages (Richardson 1999); and (4) the unconstrained model (UC) which resembles de Beer's (1930) predictions of randomness of evolutionary changes among developmental stages. Most variation tests for these models relied on polynomial adjustment: linear with a positive slope for EC, convex for HG, concave for AP, and no significant relation for UC. The same patterns for the different models are also thought to be reproduced using the phylogenetic metrics used here instead of direct estimates of variation. This is because highly conserved partitions (evolving at a rate slower than the speciation or cladogenic rate) will tend to have lower values, whereas less conserved and neutral partitions will tend to have higher values. For all indices, polynomial adjustments were conducted using model fit option in PAST software package (Hammer et al. 2001): first order polynomial (i.e., linear) for EC and second order polynomial for HG and AP. Chi-squared values were estimated for fitting errors in first- and second-order polynomial adjustments. Larger Chi-squared values mean poorer fit. Moreover, to measure the impact of the taxonomic level on pattern, analyses were also conducted on two levels: the *melanogaster* species subgroup level (four nodes), and the subgeneric level (three nodes) comprising the subgenus *Drosophila sensu stricto* and the two radiations of the subgenus *Sophophora*.

RESULTS

Figure 1 shows the TET phylogeny. When concatenated sequences of each developmental class (2nd level partitioning) were analyzed separately the same topology was recovered (data not shown). However, when phylogeny was inferred for each gene separately (1st level partitioning) some nodes were less recovered than others. This is shown in Fig. 1 by the RCFI distributions below each node. RCFI distributions ranged from 61% in the *virilis-mojavensis* (node X) and *yakuba-erecta* (node V) clades to 100% in the *pseudoobscura-persimilis* (node VIII) clades. The *grimshawi-virilis-mojavensis* clade (node IX) encompasses all species of the *Drosophila s.s.* subgenus. In some instances (four out of the 51 genes), this clade appeared sister to the Old World radiation of the subgenus *Sophophora* (node II), that is, *Drosophila willistoni* the only representative of New World *Sophophora* appearing as the earliest branching lineage. Indeed, the relationship between Old World and New World sophophorans have long been problematic (O'Grady and Kidwell 2002), but whole sequences of nuclear (*Drosophila* 12 Genomes Consortium 2007; Rosenfeld et al. 2008) and mitochondrial genomes (O'Grady and DeSalle 2008) supported their monophyly. The 100% recovery of a monophyletic *pseudoobscura-persimilis* clade (node VIII) is not surprising as these two reproductively isolated species are so closely related that they can hardly be distinguished on morphological, karyological, or even molecular basis (Machado and Hey 2003). The instability of the *yakuba-erecta* clade (node V) has previously been recognized in whole-genome analyses and was attributed to either incomplete lineage sorting (Pollard et al. 2006) or long branch attraction (DeSalle et al. in prep.). In sum, nine out of the 10 nodes of TET have not been recovered by all genes.

We analyzed the pattern of RCFI distributions among developmental stages (Fig. 2). Overall RCFI showed a significantly negative trend ($y = -0.041\ x + 0.906$; $P < 0.001$). We investigated the effect of the taxonomic level on RCFI distribution and found two very distinctive patterns. At the *melanogaster* species subgroup level (nodes IV–VII), the trend took a convex function (polynomial adjustment fit: $y = 1.11\ x^2 - 0.26\ x + 0.03$; $P < 0.05$) with its minimal extreme being in the segment polarity stage (RCFI = 55%). At the subgeneric level (nodes I, II, and IX for *Drosophila s.s.*, Old World sophophorans and the position of *D. willistoni*), the opposite trend was found as RCFI took the shape of a concave function ($y = 0.62\ x^2 + 0.22\ x - 0.04$; $P < 0.01$) with its maximal extreme being in the pair-rule stage (RCFI = 96%). Polynomial adjustment fit is in concordance with a hypothesis of an effect of Raff's hourglass hypothesis on phylogenetic signal, although the reason for the opposing trends at the two taxonomic levels remains unclear.

Figure 3 shows the distribution of BS indices among developmental stages. All BS values were positive indicating support to the TET topology. Unlike RCFI, BS patterns did not differ between global, species subgroup, and subgeneric levels as all showed convex functions ($P = 0.018$ for global and species subgroup levels and $P = 0.077$ for the subgeneric level) with minimal extreme being at the segment polarity stage. The same patterns with the same significance levels were also obtained for PBS and HPBS indices (not shown). In sum, maternal genes had higher support to the TET phylogeny than zygotic genes; among the latter, segment polarity genes had the least support to TET regardless to the taxonomic level in complete concordance with the hourglass hypothesis.

Table 1 shows the distribution of genes with significant LILD scores on the TET phylogeny (Fig. 1). Considering developmental stages (2nd level partitioning), maternal and gap partitions did not show significant incongruence, though some of their genes where incongruent at some particular nodes. Pair-rule partition was incongruent at the *yakuba-erecta* clade, and hence were segment polarity and identity partitions at the *virilis-mojavensis* clade. Considering individual genes (1st level partitioning), 11 out of the 51 genes were at odd with the TET topology of the *melanogaster* species subgroup. Only two of them belong to the maternal stage.

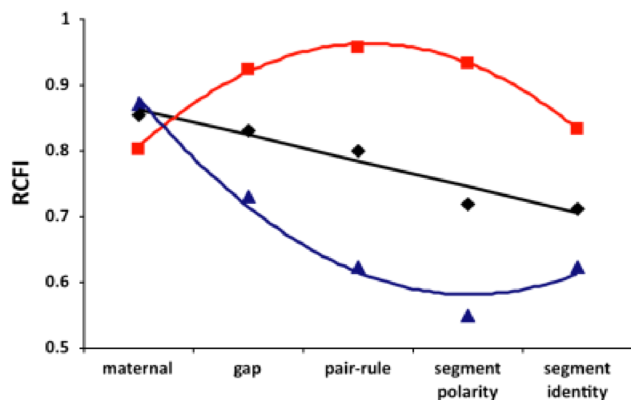

Fig. 2. Polynomial adjustment of relative consensus fork index (RCFI) among developmental stages for all nodes (black), at the *melanogaster* species subgroup level (in blue) and the subgeneric level (in red).

Table 1. List of genes with significant localized incongruent length difference (LILD) scores, showing their chromosomal position (according to the *Drosophila melanogaster* genome) and nodes where their incongruence is located (cf. Fig. 1)

Stage	Gene	Chromosome	Location	Nodes
Maternal	*armi*	3L	63E	III
	psq	2R	47A	VI
	spir	2L	38C	V
	swa	X	5F	II
	vls	2L	38B	X
Gap	*cad*	2L	38B	V
	croc	3L	78D	II, X
	ems	3R	88A	III, V, VI
	kr	2R	60F	III
Pair-rule	*h*	3L	66D	III
	odd	2L	24A	V
	ten-m	3l	79D–E	VII
Segment polarity	*en*	2R	47F	V
	inv	2R	47F	X
Segment identity	*abd-A*	3R	89A	V, VI, X
	ftz	3R	84A	VI
	lab	3R	84A	V, X
	pb	3R	84A	X
	scr	3R	84A	II, III,
	ubx	3R	89D	V

The remaining nine genes belong to the zygotic stages: gap (two genes), pair-rule (two genes), segment polarity (one gene), and segment identity (four genes). At the subgeneric level, only the node supporting the monophyly of Old World sophophorans had three incongruent genes, each belonging to a different stage: maternal, gap and segment identity. In sum, no significant pattern of LILD scores among developmental stages was found showing that different stages do not support different phylogenetic histories.

Curiously, the chromosomal distribution of the 19 incongruent genes was not random. Among the four incongruent genes located on chromosome 2L three were found at region 38B-E of *D. melanogaster*, whereas among the four incongruent genes on chromosome 2R three were found at region 47A-F. Among the seven incongruent genes on chromosome 3, four were located at region 84A (the *antennipedia* gene complex), and three at region 88A–89D (*empty spiracles* and two genes of the *ultrabithorax* complex). These regions are known to be rearrangement breakpoints in the *melanogaster* species subgroup (Lemeunier and Ashburner 1976), suggesting a major role of synteny to explain the nonrandom pattern of LILD distribution.

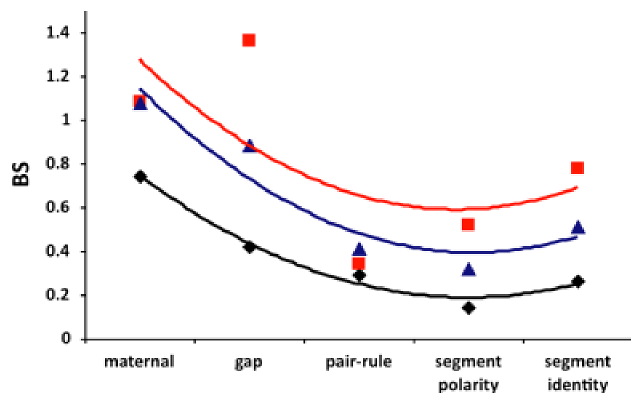

Fig. 3. Polynomial adjustment of branch support (BS) index among developmental stages for all nodes (black, $P<0.05$), at the *melanogaster* species subgroup level (in blue, $P<0.05$) and the subgeneric level (in red, $P=0.08$).

Finally, we estimated the synapomoprhy content of each partition ad hoc on the TET phylogeny. Figure 4 shows the distribution of the recapitulatory index (number of synapomorphies divided by the number of phylogenetically informative characters) among the developmental stages. Again, a convex function is resolved ($y=0.389\ x^2-0.082\ x+0.012$; $P<0.05$) with a minimal extreme lying between pair-rule and segment polarity genes. The phylogenetic content of mid-developmental stages is lower than in early and late stages, in concordance with the predictions from the hourglass model.

Pleiotropic genes, that is, those that are also expressed in larvae and pupae, may convolute the results of our study aiming at investigating the phylogenetic stages of different developmental stages. However, excluding them has not been possible, simply because it would imply the exclusion of entire classes (namely, the two last stages of segment polarity and segment identity). Among the three first stages, only a few genes were pleiotropic but they did not show any significantly aberrant values for the phylogenetic tests used in this study. Cruickshank and Wade (2008) also tested pleiotropy

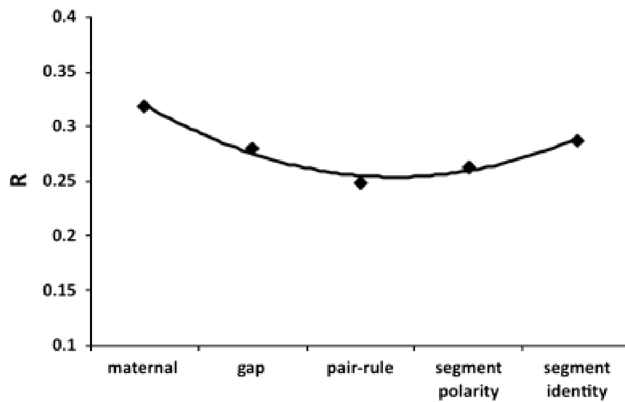

Fig. 4. Polynomial adjustment of the recapitulatory index (R) among developmental stages.

by considering the number of embryonic stages a gene is expressed in and found only a suggestive but not significant variation. Moreover, McDonal-Kreitman test of the DNA sequences of these genes did not find a significant excess of nonsynonymous mutations at any of the five developmental classes including the two last pleiotropic one. This was also true for the LILD test conducted here, which was more correlated to chromosomal location than to developmental class. For these reasons, pleiotropic genes are not thought to having skewed the results of our study.

DISCUSSION

Developmental constraints on phylogeny inference

In evo-devo, both developmental and phylogenetic constraints on character evolution occupy an important place. The major result of this study is elucidating that the resolution of phylogeny inference, crucial in evo-devo studies, is correlated to the evolutionary model of development itself. Several models have been previously developed to illustrate how development is evolutionarily self-constrained. The two most frequent models are early conservation (von Baer 1828) and the hourglass (Raff 1996) models. Support for the hourglass model in *Drosophila* segmentation genes came from comparisons of DNA sequences between close and distant species pairs of the subgenus *Sophophora* (Davis et al. 2005; Cruickshank and Wade 2008). However, these studies described the differences in the variability of each developmental stage but did not predict their differences in the phylogenetic information content due to both taxonomic limitations (only two species) and phenetic approaches (not excluding homoplasies).

Within a parsimony framework, four phylogenetic indices were analyzed: RCFI, BS and derived indices, LILD, and R. RCFI is a measure of phylogenetic resolution, whereas BS and R are measures of homology content. Polynomial adjustment of BS and R indices resulted in a convex parabola, which is in support for the hourglass model. In other words, this means that different developmental stages differ in their phylogenetic signal with mid stages being the least phylogenetically informative. A constant result was the higher signal of maternal genes than zygotic ones. However, both indices are highly sensible to the size of the data set (i.e., number of amino acid residuals or sequence length). Indeed the developmental partitions differed significantly in both their gene numbers and sequence lengths and one would expect their contribution to the stability of the TET to differ accordingly. Indeed about 45% of the concatenated matrix belong to the maternal genes partition. Nonetheless, even when the four zygotic partitions are pooled together their BS and R indices are still significantly inferior to those of maternal genes.

Selective constraints and phylogenetic incongruence

Significant LILD scores of a particular partition indicate that it is supporting a different evolutionary history for a particular clade than the one predicted by the TET. That is why they have been used to detect horizontal gene transfers, introgression, co-evolution, recombination, and convergent evolution (Thornton and DeSalle 2000; Narechania et al. 2005; Ramirez 2006). Nineteen out of the 51 genes analyzed here had significant LILD scores, some of which at up to three nodes. Because all of these genes are found in the nuclear genome, their significant incongruence may arise from convergent selection or chromosomal location.

Convergent selection would be expected if incongruent genes significantly belonged to different stages. Cruickshank and Wade (2008) conducted McDonald–Kreitman tests of these genes between the homosequential species *D. melanogaster* and *D. simulans* using several strains of *D. simulans*. With the exception of pair-rule genes, all tests gave nonsignificance results indicating that selection might be substantial. The higher levels of polymorphism and divergence of maternal genes relative to those of zygotic ones are due to relaxed selection on genes expressed in a single sex, that is, copies of maternal genes in males are hidden from natural selection resulting in higher levels of nucleotide variation at replacement sites within populations (Wade 1998). We have also conducted relative rate tests (Tajima 1993) and found that the last two stages and the last three stages gave significant scores of rate heterogeneity at the subgeneric and the species subgroup levels, respectively (not shown). This indicates that earlier stages evolve under a more quasi-neutral pattern as predicted by the relaxed selection hypothesis.

LILD scores were distributed randomly across stages in contrast to convergent selection predictions. Nonetheless,

genes of the segment identity (hox) stage present an important exception. Six out of these eight genes were significantly incongruent. Although this may be attributed to a convergent directional selection on this last stage, segment identity genes are all clustered on the left arm of chromosome 3, and thus their phylogenetic incongruence may be an artifact of synteny at this region. Synteny is expected to play a major role in hox gene evolution because chromatin looping and chromosomal colinearity are essential in the regulation of their transcription (Fraser and Bickmore 2007; Montavon et al. 2008).

Protein sequences of genes located at rearrangement breakpoints in *Drosophila* were recently shown to change faster than those located within the rearranged fragments (Ranz et al. 2007). Moreover, the number of incongruent genes is evenly distributed between the two arms of chromosome 2, and the greatest number of incongruent genes is found on the left arm of chromosome 3. This is in concordance with predictions on the effect of rearrangement regions in the *melanogaster* species subgroup where the greatest distribution of incongruence was found here on one major pericentric inversion on chromosome 2 and one major paracentric inversion on chromosome 3L occurred during the evolution of the subgroup (Lemeunier and Ashburner 1976; Schaeffer et al. 2008). These findings highlight the fact that in spite of differences in conservation among different developmental stages linkage constraints might play a more important role in shaping the phylogenetic history of some genes than functional constraints.

Rate dependence for topological resolution

Unlike the previous character-based measurements, RCFI scores were not estimated from a simultaneous analysis. Instead, a taxonomic congruence approach was followed wherein phylogenetic analyses are conducted one by one and each tree is compared with the TET. This individual gene approach makes any discrepancy among genes more influenced by sequence length (i.e., no hidden support to TET can be revealed by interaction with other genes as in a simultaneous analysis) and rate heterogeneity, resulting in more frequent artifacts of long branch attraction and random rooting (Bergsten 2005). Although interpretations of BS and R indices in terms of developmental constraints and LILD scores in terms of linkage constraints were straightforward, interpretation of the RCFI was not. The negative slope of overall RCFI scores contradicts quantitative predictions of all known developmental models (Poe and Wake 2004). More interestingly are the completely inverse parabolas obtained for RCFI scores when different taxonomic levels are considered.

The contrasting patterns at the two taxonomic levels can be explained in light of the heterogeneity in the rate of evolution of proteins expressed at different developmental stages as predicted from the hourglass model. At the *melanogaster* subgroup level originating around 10–15 Ma (Russo et al. 1995; Tamura et al. 2004), maternal genes evolving under relaxed selection as suggested by Cruickshank and Wade (2008) resolve better phylogenetic relationships. At the subgeneric level originating around 40–65 Ma (Russo et al. 1995; Tamura et al. 2004), maternal genes became substantially homoplasic that slowly evolving zygotic genes resolve better the phylogeny.

Differences in the rate of evolution among the developmental stages on the shape of the developmental function may explain the conflicting results surrounding the presence of an hourglass model in Metazoa. The hourglass model was found between *D. melanogaster* and *D. simulans* (2–4 Ma; Cruickshank and Wade 2008), species of the *melanogaster* subgroup (10–15 Ma; this study), and between *D. melanogaster* and *D. pseudoobscura* (20–40 Ma; Davis et al. 2005), but was not found between the two subgenera (40–65 Ma; this study) where a concave parabola supporting an adaptive penetrance model was observed. The direct quantification of the hourglass pattern may also be convoluted at higher taxonomic levels when othology determination will be more difficult. For example, the *bicoid*-driven maternal regulation system is derived in cyclorrhaphan Diptera and the genetic basis of maternal regulation in other arthropods is still unknown (Peel et al. 2005) and one may suspect an early conservation model at an arthropod scale. Similar discrepancy was observed in analysis of developmental protein sequences in Vertebrates: support for the hourglass model in comparing human and mouse (65–85 Ma; Hazkani-Covo et al. 2005), for the adaptive penetrance model in eutherian mammals (97–110 Ma; Hanada et al. 2007), and for the early conservation model in comparing zebrafish and mouse (400 Ma; Roux and Robinson-Rechavi 2008).

In conclusion, our analysis gives insight to the relation between ontogeny and phylogeny in *Drosophila* from a character-based approach. Although support for an hourglass model comes from the different homology content of the different developmental stages as revealed by BS and R indices, the finding of an hourglass using individual gene analysis highly depends on the taxonomic level being studied (as in RCFI). This finding would certainly be further tested by two means. First, as genomic data are keeping growing, the future inclusion of additional taxonomic levels in phylogenetic analyses of developmental genes would certainly increase our knowledge about how development constrains evolution and vice-versa. And second, by further evo-devo analyses investigating the functional developmental divergence in the 12 *Drosophila* species. We hope that the pattern deduced here for sequence evolution of developmental proteins can serve as a testable hypothesis for the correlation between sequence and functional divergence in future evo-devo investigations of body segmentation in *Drosophila* and other closely related species.

Acknowledgments

The authors are very grateful to two anonymous reviewers for their constructive criticisms. This work was funded by a research price from the Foundation Bettencourt-Schueller and a post-doctoral fellowship from the Lounsbery Foundation for A. Yassin.

REFERENCES

Baker, R. H., and DeSalle, R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst. Biol.* 46: 654–673.

Bergsten, J. 2005. A review of long-branch attraction. *Cladistics* 21: 163–193.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42: 795–803.

Brody, T. 1999. The interactive fly: gene networks, development and the internet. *Trends Genet.* 15: 333–334.

Cruickshank, T., and Wade, M. J. 2008. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in *Drosophila. Evo. & Devo.* 10: 583–590.

Colless, D. H. 1980. Congruence between morphometric and allozyme data for *Menidia* species: a reappraisal. *Syst. Zool.* 29: 288–299.

Davis, J. C., Brandman, O., and Petrov, D. A. 2005. Protein evolution in the context of *Drosophila* development. *J. Mol. Evol.* 60: 774–785.

de Beer, G. R. 1930. *Embryology and Evolution*. Clarendon, Oxford.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.

Farris, J. 1989. The retention index and the rescaled constancy index. *Cladistics* 5: 417–419.

Farris, J. S., Kallersjö, M., Kluge, A. G., and Bult, C. 1995. Constructing a significance test for incongruence. *Syst. Biol.* 44: 570–572.

Fraser, P., and Bickmore, W. 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature* 447: 413–417.

Gatesy, J., O'Grady, P., and Baker, R. 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics* 15: 271–313.

Gould, S. J. 1977. *Ontogeny and Phylogeny*. Belknap, Cambridge, MA.

Haeckel, E. 1866. *Generelle morphologie der organismen. Allgemeine grundzüge der organischen formen-wissenschaft, mechbanisch begründet durch die von Charles Darwin reformire descendenz-theorie*. Georg Reimer, Berlin.

Hammer, Ø., Harper, D. A. T., and Ryan, P. D. 2001. PAST: paleontological software package for educational data analysis. *Paleontol. Elect.* 4: 1–9.

Hanada, K., Shiu, S.-H., and Li, W.-H. 2007. The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Mol. Biol. Evol.* 24: 2235–2241.

Hazkani-Covo, E., Wool, D., and Graur, D. 2005. In search of the vertebrate phylotypic stage: a molecular examination of the developmental hourglass model and von Baer's third law. *J. Exp. Zool.* 304B: 150–158.

Kluge, A. G., and Farris, J. S. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 40: 257–270.

Kumar, S., Dudley, J., Nei, M., and Tamura, K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinformat.* 9: 299–306.

Lemeunier, F., and Ashburner, M. 1976. Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. *Proc. R. Soc. Lond. B.* 193: 275–294.

Machado, C. A., and Hey, J. 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. Lond. B.* 270: 1193–1202.

Miyamoto, M., and Fitch, W. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Syst. Zool.* 35: 230–240.

Montavon, T., Le Garrec, J.-F., Kerszberg, M., and Duboule, D. 2008. Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev.* 22: 346–359.

Narechania, A., Chen, Z., DeSalle, R., and Burk, R. D. 2005. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J. Virol.* 79: 15503–15510.

O'Grady, P., and DeSalle, R. 2008. Out of Hawaii: the origin and biogeography of the genus *Scaptomyza* (Diptera: Drosophilidae). *Biol. Lett.* 4: 195–199.

O'Grady, P. M., and Kidwell, M. G. 2002. Phylogeny of the subgenus *Sophophora* (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. *Mol. Phyl. Evol.* 22: 442–453.

Peel, A. D., Chipman, A. D., and Akam, M. 2005. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat. Rev. Genet.* 6: 905–916.

Poe, S., and Wake, M. W. 2004. Quantitative tests of general models for the evolution of development. *Am. Nat.* 164: 415–422.

Pollard, D. A., Iyer, V. N., Moses, A. M., and Eisen, M. B. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genet.* 2: 1634–1647.

Raff, R. A. 1996. *The Shape of Life: Genes, Development, and the Evolution*. John Wiley and Sons, New York.

Ramirez, M. J. 2006. Further problems with the incongruence length difference test: "hypercongruence" effect and multiple comparisons. *Cladistics* 89: 289–295.

Ranz, J. M., et al. 2007. Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol.* 5: 1366–1381.

Richardson, M. K. 1999. Vertebrate evolution: the developmental origins of adult variation. *Bioessays* 21: 604–613.

Rosenfeld, J. R., DeSalle, R., Lee, E., and O'Grady, P. M. 2008. Using whole genome presence/absence data to untangle function in 12 Drosophila genomes. *Fly* 2: 291–299.

Roux, J., and Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. *PLoS Genet.* 4: e1000311.

Russo, C. A. M., Takezaki, N., and Nei, M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* 12: 391–404.

Sander, K. 1983. The evolution of patterning mechanisms: gleanings from insect embryogenesis and spermatogenesis. In B. C. Goodwin, N. Holder, and C. C. Wyle (eds.). *Development and Evolution*. Cambridge University Press, Cambridge, UK, pp. 137–159.

Schaeffer, S. W., et al. 2008. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* 179: 1601–1655.

Swofford, D. L. 2003. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sinauer Associates, Sunderland, MA.

Tamura, K., Subramanian, S., and Kumar, S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol. Biol. Evol.* 21: 36–44.

Tajima, F. 1993. Simple methods for testing molecular clock hypothesis. *Genetics* 135: 599–607.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. ClustalW improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22: 4673–4680.

Thornton, J. W., and DeSalle, R. 2000. A new method to localize and test the significance of incongruence: detecting domain shuffling in the nuclear receptor superfamily. *Syst. Biol.* 49: 183–201.

Tweedie, S., et al. 2009. FlyBase: enhancing *Drosophila* gene ontology annotations. *Nuc. Acids Res.* 37: D555–D559.

von Baer, K. E. 1828. *Entwicklungsgeschichte der Thiere, Beobachtung und Reflexion*. Bornträger, Königsberg.

Wade, M. J. 1998. The evolutionary genetics of maternal effects. In T. Mousseau and C. Fox (eds.). *Maternal Effects as Adaptations*. Oxford University Press, Misc: pp. 5–21.