

Comparative Analysis of Chromosome Counts Infers Three Paleopolyploidies in the Mollusca

Nathaniel M. Hallinan* and David R. Lindberg

Department of Integrative Biology, University of California Berkeley

*Corresponding author: E-mail: nmhallinan@gmail.com.

Accepted: 8 August 2011

Abstract

The study of paleopolyploidies requires the comparison of multiple whole genome sequences. If the branches of a phylogeny on which a whole-genome duplication (WGD) occurred could be identified before genome sequencing, taxa could be selected that provided a better assessment of that genome duplication. Here, we describe a likelihood model in which the number of chromosomes in a genome evolves according to a Markov process with one rate of chromosome duplication and loss that is proportional to the number of chromosomes in the genome and another stochastic rate at which every chromosome in the genome could duplicate in a single event. We compare the maximum likelihoods of a model in which the genome duplication rate varies to one in which it is fixed at zero using the Akaike information criterion, to determine if a model with WGDs is a good fit for the data. Once it has been determined that the data does fit the WGD model, we infer the phylogenetic position of paleopolyploidies by calculating the posterior probability that a WGD occurred on each branch of the taxon tree. Here, we apply this model to a molluscan tree represented by 124 taxa and infer three putative WGD events. In the Gastropoda, we identify a single branch within the Hypsogastropoda and one of two branches at the base of the Stylommatophora. We also identify one or two branches near the base of the Cephalopoda.

Key words: whole genome duplication, Mollusca evolution, birth–death process, karyotype evolution.

Introduction

Polyploidy has long been recognized as an important mechanism of genetic evolution (Taylor and Raes 2004), and over the last decade, as full genome sequences have become available, a great deal of research has been invested into the analysis of whole-genome duplications (WGDs) (e.g., Byrne and Blanc 2006; Semon and Wolfe 2007). Polyploid species and individuals are common in plants (Wendel 2000); evidence is accumulating that WGDs also occur in the opisthokonts and have played an important role in the evolution of their genomes (McLysaght et al. 2002; Vandepoele et al. 2004; Wolfe 2004). In contrast to other modes of genome evolution, polyploidy events affect the entire genome at once. By duplicating every gene in the genome, a large amount of redundant genetic information is created, which can be used as raw material for evolutionary innovations (Haldane 1932; Ohno 1967). It has been suggested that in several cases the modification of this raw material has been important in the evolution of key innovations, such as glucose fermentation in yeast (Piskur

2001) and the immune system of vertebrates (Kasahara 2007).

Despite the large effort put into the analysis of genome duplications, the identification and confirmation of such duplications, especially ancient ones, has proved problematic. To conclusively demonstrate a paleopolyploidy event, several complete genomes must be sequenced both from taxa that are directly descended from the original polyploid individual and from their relatives that diverged shortly before the WGD (Wong et al. 2002; Woods et al. 2005). Even to identify likely cases of paleopolyploidy large numbers of genes must be sequenced in several closely related taxa (Spring 1997; Blanc and Wolfe 2004), and although sequencing costs continue to decline, it remains costly in terms of laboratory time and materials (Ansorge 2009).

The comparison of lineages that diverged shortly before and after the WGD is critical to the accurate reconstruction of the event (Scannell et al. 2007) (fig. 1). Thus, it would be beneficial to identify putative positions of a WGD before initiating research into the effects of paleopolyploidy. Here, we

The Author(s) 2011. Published by Oxford University Press on behalf of the *Society for Molecular Biology and Evolution*.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

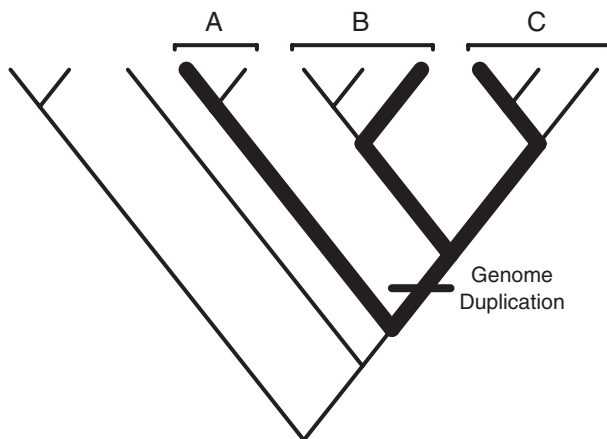


FIG. 1.—An ideal sampling of three taxa for investigating the genome duplication indicated by the dash. By selecting one taxon each from clades A, B, and C, the investigator would minimize the amount of shared history either before or after the event and thus have the best chance of accurately reconstructing the effects of the duplication on the genome.

show that by using karyotypic data we are able to increase taxon sampling to accurately identify the branches on which the duplication occurred and thus guide the selection of taxa for genome sequencing.

Although it is possible to identify clades within a phylogeny that have approximately twice the number of chromosomes as their close relatives, such distributions of chromosome numbers may be the consequence of a WGD or of ordinary processes of aneuploidy. In order to distinguish the background rate of change in chromosome number from doublings caused by WGDs, we used a likelihood method in which the background rate of chromosome number evolution is modeled as a birth–death process. The birth–death process is a stochastic model often used to compare the numbers of genes in gene families between organisms (Lynch and Conery 2003; Hahn et al. 2005; Novozhilov et al. 2006). Under this process, the rate of chromosome duplication or loss is proportional to the number of chromosomes in the genome. This seems an appropriate model as aneuploidy is usually a consequence of nondisjunction; thus both gains and losses are likely to occur at approximately equal rates for any random chromosome.

We used a model in which chromosome number evolved by the birth–death process alone as a null model and compared it with a duplication model in which the number of chromosomes could also double in a single stochastic event. The maximum likelihoods for a set of chromosome counts were calculated under each model and compared with simulations in order to determine whether the birth–death process alone was a reasonable fit for the data, whether the duplication model was a significantly better fit than the birth–death process alone, and whether the birth–death process was a sufficient approximation of the background

rate of chromosome number evolution in the presence of WGDs. Furthermore, the parameters inferred under the duplication model were used to calculate the posterior probability that a duplication occurred on each branch of the tree. Mayrose et al. (2010) recently published a similar likelihood method. They used models in which the rates of chromosome increase and decrease were independent of the number of chromosomes and models in which those rates were linearly related to the number of chromosomes; the birth–death process is a special case of the latter model.

Here, we analyze the Mollusca for the occurrence of WGDs. The Mollusca is a large and disparate clade with members that play an important role in marine and terrestrial ecosystems. Furthermore, they are one of the most diverse groups within the Lophotrochozoa, the least studied of the three major clades of bilaterian animals. Currently, genomics studies in Mollusca are still preliminary, with fewer than five genomes sequenced or in progress (Chapman et al. 2007). Natural polyploid species of molluscs have been recognized for decades (e.g., Patterson 1969; Goldman et al. 1983), and it has been suggested that the Stylommatophora (Steusloff 1942) and the Neogastropoda (Hinegardner 1974) are polyploid. By determining which molluscan clades are polyploid, we can identify potential taxa for investigation and help to guide future research into molluscan genomics.

Materials and Methods

Phylogeny

No single phylogenetic study has included all the taxa in our analysis, so we used phylogenies from several sources to assemble a tree for the Mollusca and then pruned taxa for which we did not have chromosome data. Although a tree constructed from a single analysis would have been preferred, such an analysis is beyond the scope of this study. We therefore used a reverse compartmentalization method to generate the tree used here (Mishler 1994; Nixon et al. 1994). We did this by identifying published trees that represented the terminal taxa in Haszprunar's (2000) molluscan phylogeny and in Ponder and Lindberg's (1997) gastropod phylogeny. We assumed that each of these terminal taxa represented a compartmentalized (or collapsed) monophyletic tree and therefore by placing the published trees at each of the respective tips we were simply expanding the terminal taxon. Mishler (1994) has shown this method is robust in representing diverse yet clearly monophyletic clades in larger scale cladistic analyses, and we simply reversed this approach by expanding rather than collapsing each terminal clade using a published phylogeny. Trees for the Polyplacophora (Okusu et al. 2003), Bivalvia (Giribet and Wheeler 2002), and Cephalopoda (Lindgren et al. 2004) were then placed on the Haszprunar backbone. Within the gastropod tree, Barker (2001) was used for the phylogeny within the Heterobranchia and Wade et al. (2006) for more detailed

relationships among the pulmonate families. Wägele et al. (2008) provided branching patterns within the Opisthobranchia, and Colgan et al. (2007) provided placement for additional families not included by Wägele et al. (2008). Pruning of trees in our reconstruction was done when chromosomal data for the taxa included in the original phylogenetic analyses were not available, as the method we employed to calculate WGDs required chromosomal data for every taxon. The presence of additional data may affect the results, as it may for any scientific analysis.

All likelihood models require branch lengths to estimate the probability of a transition along a given branch, and any rates used in such models are proportional to the units of branch length. Therefore, we assumed that under a null model the rates of chromosome duplication and loss were constant with respect to time and therefore the branch lengths for the tree should be in years between speciation events. We deduced the timing of each node in our tree by identifying the first occurrence for each of the terminal taxa in our study as well as for larger clades containing one or more of these taxa. These dates were determined from the Paleobiology Database (www.paleodb.org) with the following exceptions: Nishiguchi and Mapes (2008) provided first occurrence dates for the cephalopods, and Solem and Yochelson (1979) and Zilch (1959–1960) were used for the pulmonate gastropods. Every node was fixed at the oldest date for the first appearance of any of its descendants. This method resulted in several branches of length zero.

We could not assign a length to a terminal branch if it led to one of two sister taxa with no fossil occurrence nor to an internal branch if one of the two clades immediately descended from it had a longer fossil history than the internal branch's sister clade. These discrepancies likely resulted from the vagaries of the fossil record and do not represent the actual time between lineage splitting events. Therefore, we assumed that each node had preceded both its daughter nodes by at least some minimum number of years. Three trees were constructed with zero length branches expanded to 10^4 , 10^5 , and 10^6 years, and we will refer to these as Tree- 10^4 , Tree- 10^5 , and Tree- 10^6 , respectively.

Chromosome Number

Karyotypic data and gastropods were mined from the comprehensive reviews of Patterson (1969), Patterson and Burch (1978), Nakamura (1986), and Thiriot-Quievreux (2003). We derived karyotype data for the bivalves from Patterson (1969), Nakamura (1985), and Thiriot-Quievreux (2002). Karyotype data sources for the Cephalopoda included Nakamura (1985), Gao and Natsukari (1990), and Vitturi et al. (1990). Polyplacophoran data was compiled from Odierna et al. (2008) and chromosome numbers for the Scaphopoda from Ieyama (1993). When karyotype data for multiple

species within a terminal taxon was found, we used the mode for that taxon (White 1973). If there were multiple modes, we used the mode closest to the median, and if there were two modes equally distant from the median, we chose one at random.

Phylogenetic Signal

To test for phylogenetic signal, we randomized the karyotype data among the tips 9,999 times and calculated the number of steps for the randomized data sets using ordered parsimony and a parsimony model for which the cost for going from m chromosomes to n chromosomes along a branch is the absolute value of $\ln(n) - \ln(m)$. Under this second parsimony model, the cost of adding or losing chromosomes is proportional to the number of chromosomes in the genome (as it is under the birth-death process) but it still retains the quality of an ordered parsimony model because the cost of going from state x to state z equals the cost of going from state x to state y plus the cost of going from state y to state z , when y is intermediate in value between x and z . We then compared the number of steps for our randomized data sets to the number of steps for our actual data. Because there should be fewer steps between closely related individuals than distantly related ones, there will be fewer steps under either parsimony model for a data set with high phylogenetic signal compared with a random data set.

Likelihood Model

For our null hypothesis, we assumed that each chromosome has an equal probability of duplicating or splitting and an equal probability of being lost at any time. Thus, there is a constant duplication rate, λ , and loss rate, μ , for each chromosome in the genome. Hallinan (2011) showed how to calculate the probability that m_i chromosomes at the beginning of branch i of the taxon tree will leave N_i chromosomes at the end of that branch, assuming that all m_i chromosomes survive to the end of that branch.

$$P_1(N_i|m_i) = \binom{N_i - 1}{m_i - 1} (1 - u_i)^{m_i} u_i^{N_i - m_i}, \quad (1)$$

where

$$u_i \equiv \begin{cases} \frac{\exp((\lambda - \mu)t_i) - 1}{\exp((\lambda - \mu)t_i) - a} & \text{if } \lambda \neq \mu \\ \frac{\lambda t_i}{\lambda t_i + 1} & \text{if } \lambda = \mu \end{cases}$$

t_i is the length of branch i and $a = \mu/\lambda$. We also know from Kendall (1948) that if E_i is the probability that a chromosome present at the base of branch i is lost by the end of that branch, then $E_i = au_i$.

Using this result, we can now calculate the probability that M_i chromosomes at the beginning of the branch leave N_i chromosomes at the end of branch i , allowing for the

possibility that any of the initial chromosomes could be lost. In this case, m_i is the number of chromosomes in the initial group of M_i chromosomes that give rise to the N_i chromosomes at the end of the process. The other $M_i - m_i$ chromosomes are lost. There are $\binom{M_i}{m_i}$ different ways to arrange $M_i - m_i$ lost chromosomes among M_i initial chromosomes. Therefore, we can calculate this probability by summing over all the possible numbers of surviving chromosomes.

$$P_2(N_i|M_i) = \sum_{m_i=1}^{\min(M_i,N_i)} \binom{M_i}{m_i} E_i^{M_i-m_i} P_1(N_i|m_i) \quad (2)$$

$$= \sum_{m_i=1}^{\min(M_i,N_i)} \binom{M_i}{m_i} \binom{N_i-1}{m_i-1} [(1-au_i)(1-u_i)]^{m_i} \times (au_i)^{M_i-m_i} u_i^{N_i-m_i}.$$

This equation is equivalent to the one given by Bailey (1964, p. 94) and Foote et al. (1999), although it takes a different form, and here, we provide an alternative derivation. We can assume that all the chromosomes were not lost on any branch of the tree, so that the transition probability between any two states M_i and N_i is the probability of going from M_i to N_i chromosomes along branch i of length t_i , given that N_i is at least one.

$$P_3(N_i|M_i) = \frac{P_2(N_i|M_i)}{1 - E_i^{M_i}}. \quad (3)$$

We used this equation to calculate the probability of a set of chromosome counts at the tips of a phylogeny conditioned on a particular value for λ , μ , and the number of chromosomes at the root. We calculated the likelihood for the entire tree by proceeding down from the tips of the tree to the root and marginalizing over all the possible states at the internal nodes (Felsenstein 1973, 1981), such that we calculated the probability of the data above branch i given M_i by summing over all the possible values of N_i .

$$P(C_i|M_i) = \sum_{N_i=1}^{\infty} p_i(N_i|M_i)P(C_i|N_i), \quad (4)$$

where $p_i(N_i|M_i)$ is the transition probability of going from M_i to N_i chromosomes and C_i is the chromosome counts on the tips above branch i . If we define i^+ and i^- as the two branches immediately descended from i , then $N_i = M_{i^+} = M_{i^-}$ and $P(C_i|N_i)$ can be calculated as $P(C_{i^+}|M_{i^+})P(C_{i^-}|M_{i^-})$, which can each in turn be calculated with equation (4). We also calculated $H_i(X)$, the highest integer such that $P(C_i|N_i = H_i(X)) > P(C_i|N_i = H_i(X) + 1)$ and $\exp(X) \times P(C_i|N_i = H_i(X)) > \sum_{N_i \in \Omega_i} P(C_i|N_i)$ where Ω_i is the set of all the values of N_i for which $P(C_i|N_i)$ has already been calculated. As there are an infinite number of possible states at each internal node, we did not calculate probabilities for values of N_i that were greater than 200 or greater than

$H_{i^+}(5)$, $H_{i^-}(5)$, and $H_i(15)$. We ran multiple tests in order to confirm that these values were set conservatively and did not result in lower likelihoods. In this case, we assumed that all evolution of chromosome numbers occurred by the birth–death process, so that $p_i(N_i|M_i) = P_3(N_i|M_i)$.

We added WGDs to this model by assuming that they occur at some constant rate, δ , and that no more than one WGD occurred on any branch. We call the number of genome duplications that happened on branch i , D_i . The transition probability of going from M_i to N_i chromosomes along a branch with no full genome duplications is simply the transition probability from the birth–death process times the probability that no genome duplication occurred.

$$P_4(N_i, D_i = 0|M_i) = \exp(-\delta t_i) P_3(N_i|M_i). \quad (5)$$

In order to calculate the transition probability of going from M_i to N_i chromosomes along a branch with one full genome duplication, we divide the branch into two discrete time periods at the time of the duplication, t_δ . In that case, the number of chromosomes before a duplication is N_δ and the time between the start of the branch and the duplication is $t_i - t_\delta$; after the duplication, there will be $2N_\delta$ chromosomes. We can calculate the transition probability by summing over all the possible values of N_δ and integrating over t_δ from zero to t_i .

$$P_4(N_i, D = 1|M_i) = \delta \exp(-\delta t_i) \times \int_0^{t_i} \sum_{N_\delta=1}^{\infty} P_3(N_\delta|M_i) P_3(N_i|2N_\delta) dt_\delta. \quad (6)$$

We calculated this integral using numerical integration, breaking each branch into sections and assuming that the duplication happened in the middle of that section. We used three sections, as we found that three sections yielded the same results as ten sections. We calculated the sum by treating the duplication event as an internal node and limiting the number of possible states as above.

Thus, we can calculate the likelihood of specific values of λ , μ , and δ conditioned on the numbers of chromosomes at the root using equation (4), except we now calculate the transition probability as $p_i(N_i|M_i) = P_4(N_i|M_i)$, where:

$$P_4(N_i|M_i) \approx P_4(N_i, D_i = 0|M_i) + P_4(N_i, D_i = 1|M_i). \quad (7)$$

This excludes the possibility of multiple duplications on any branch, which is in general very small. These calculations were performed using the program GDCN 1.0. Windows executables and code are available from <http://code.google.com/p/gdcn/>.

Model Comparison

We compared three different models by fitting the parameters of those models to our data set by maximum likelihood. The “static” model is the simplest model and

assumes that $\lambda = \mu$ and that $\delta = 0$. In order to show that our data are better explained by including WGDs, we compared the static model to a “duplication” model in which δ is allowed to take values greater than 0. It is possible that the superior fit of the duplication model is a consequence of a strong tendency for the number of chromosomes to increase rather than to paleopolyploidy events. To account for this possibility, we also compared these two models to the “trend” model that allows λ and μ to take different values but assumes that $\delta = 0$. We examined all three models on all three trees in order to assure that no set of branch lengths was unduly affecting our conclusions.

Our calculation of the likelihood is conditioned on the number of chromosomes at the root. Because we do not know the actual number of chromosomes in the common ancestor of all molluscs, we treated the ancestral chromosome count as another parameter to be set by our maximum likelihood search. In order to justify this approach, we initially calculated maximum likelihoods conditioned on a range of ancestral chromosome values from 1 to 80. By comparing a large range of ancestral mollusc chromosome counts, we could see how strongly our model supported the maximum likelihood reconstruction of the root and be certain that our model choice was not overly biased by our reconstruction of the ancestral chromosome count.

The maximum likelihoods for all models under all sets of root values were used to calculate the Akaike Information Criterion (AIC) for those models. The AIC is defined as twice the number of parameters in the model minus twice the maximum likelihood. The AICs of the different models were compared in order to select the best fit model for the data. Models that fit the data better have lower AICs (Akaike 1974).

Identifying Branches with Duplications

In order to identify branches on which it is likely that a WGD occurred, we calculated the posterior probability that there was a duplication on each branch of the tree. For each branch, i , we used the maximum likelihood values of λ , μ , δ and the ancestral chromosome count to calculate two likelihoods: $P(C_r, D_i = 1 | N_r)$ and $P(C_r, D_i = 0 | N_r)$, where branch r is the root of the tree. To make these calculations, we assume that $p_j(N_j | M_j)$ is $P_4(N_j | M_j)$ on every branch of the tree except i where $p_j(N_j | M_j)$ is $P_4(N_j, D_i = 1 | M_j)$ or $P_4(N_j, D_i = 0 | M_j)$, according to which likelihood we want to calculate. In that case, $P(C_r | N_r) \approx P(C_r, D_i = 1 | N_r) + P(C_r, D_i = 0 | N_r)$ and the posterior probability of a duplication on that branch will be $P(D_i = 1 | C_r, N_r) = P(C_r, D_i = 1 | N_r) / P(C_r | N_r)$. These values were calculated for all three trees, using the maximum likelihood values for λ , μ , δ , and the root.

Simulations and Model Fit

In addition to demonstrating which of these three models fits the data better, it is also necessary that these models

provide a reasonable description of the data themselves. In order to make this comparison, we simulated 1,000 data sets on Tree-10⁶ under the static and the duplication model using maximum likelihood parameter values. Maximum likelihoods were then calculated for each of these data sets under the models used to generate the data. These maximum likelihoods were then compared with the maximum likelihood of our data to potentially reject the hypothesis that the data were generated by this model. Failure to reject this hypothesis was taken to indicate that the model was a reasonable approximation of the actual process by which the real data were generated. The data sets simulated under the static model were also evaluated under the duplication model, and the difference between the maximum likelihoods under each model for each data set were compared with our data set in order to confirm that the duplication model was a significantly better fit for the data.

We also considered it important to demonstrate that the birth–death process was an appropriate model for the background rate of chromosome number evolution independently of where any WGDs occurred. To do this, we simulated 1,000 data sets on Tree-10⁶ using the maximum likelihood values for the root count, λ and μ calculated under the duplication model. Under these simulations, WGDs did not occur at random but instead always occurred on the branches for which our actual data showed a high posterior probability of a WGD. In order to test the effects that a duplication on a minimum length branch would have on the data set, we also simulated 1,000 data sets using maximum likelihood values for the root count, λ , μ , and δ on Tree-10⁶ and placed one duplication at random on a minimum length branch. We calculated maximum likelihoods for these simulated data sets using the duplication model and compared those likelihoods with the one calculated for our actual data in order to potentially reject the use of the birth–death process for our background rate of chromosome evolution.

Results

Phylogeny, Chromosome Counts, and Signal

The topology for the phylogeny we used is shown in figure 2a. Most traditional molluscan clades are monophyletic in this tree but both Caenogastropoda and Sigmurethra are paraphyletic. Figure 2b shows the same tree with the branch lengths scaled to match the branch lengths we derived from paleontological data. In this second tree, 68 of 122 internal branches had zero branch lengths, as a consequence of basal branching taxa in a clade having a younger fossil occurrence than taxa nested within the clade, and 6 of 124 terminal branches also had zero branch lengths because two sister taxa had no fossil occurrences. Zero length branches are shown as polytomies on the tree.

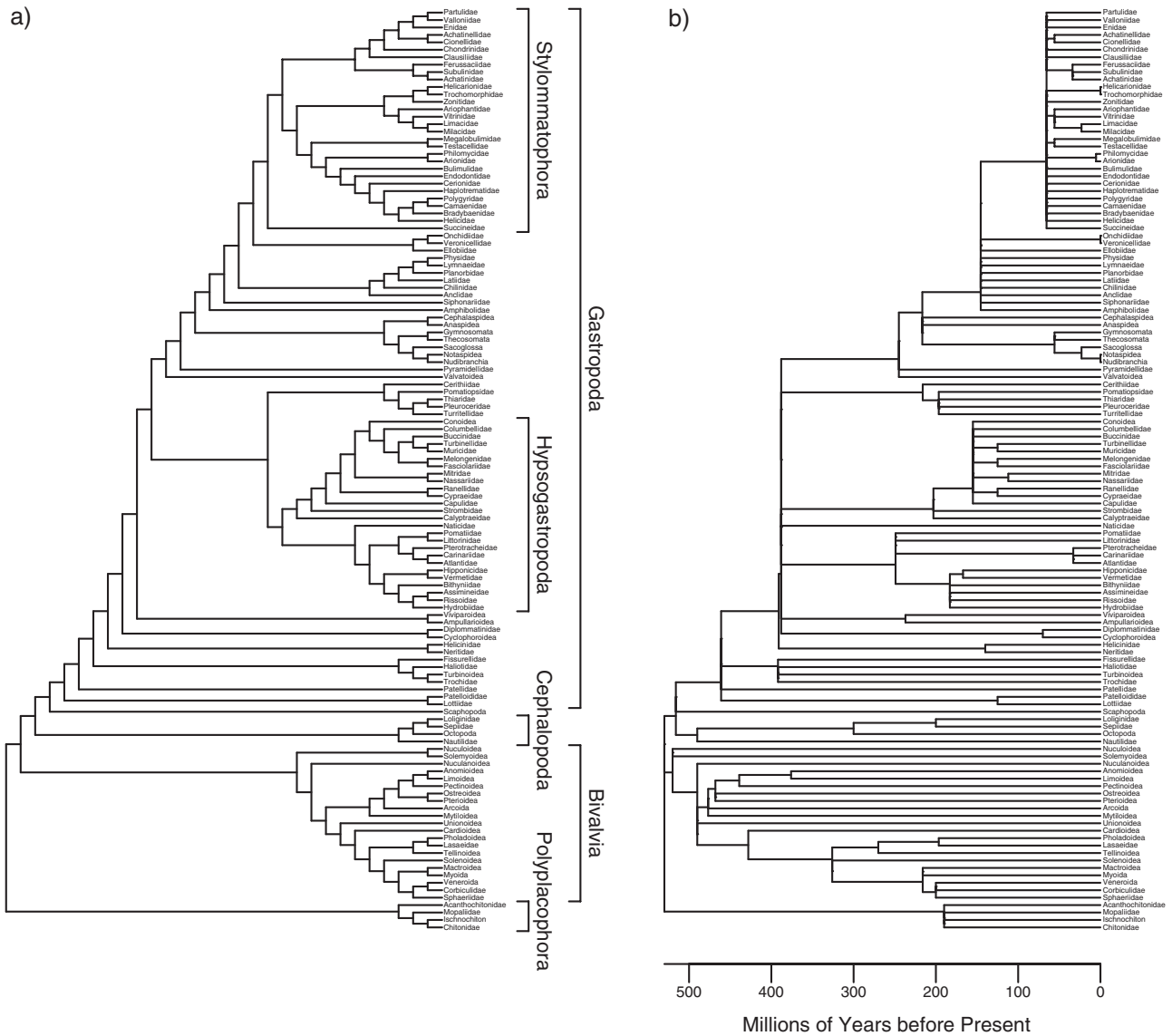


FIG. 2.—(a) The cladogram of the Mollusca used in this study. Several clades important to this study are identified with brackets. (b) The phylogeny of molluscan taxa used in this study. The topology of this tree is the same as the tree in (a), but here branch lengths are shown in millions of years as derived from the fossil record using the Paleobiology Database. Zero length branches are shown as polytomies.

Chromosome counts were obtained for 997 molluscan species within 124 terminal taxa and included members of five major extant classes (Polyplacophora, Bivalvia, Cephalopoda, Scaphopoda, and Gastropoda) (supplementary table, Supplementary Material online). Chromosomal data were unavailable for the Monoplacophora and Aplousobranchia (Chaetodermomorpha and Neomeniomorpha). When multiple values were available for a terminal taxon, the mode of these chromosome numbers was used to represent the terminal clade in our phylogenetic analysis. In most cases, all the chromosome counts in each terminal clustered around the modes, but several clades have members with highly divergent counts. For example, within the Anomioidea,

Loliginidae, Viviparoidea, Thiaridae, Ancyliidae, and Planorbidae, there are taxa with chromosome counts two times the mode for the entire clade implying recent polyploidies. In the Planorbidae genus *Bulinus*, there are species with three and even four times the mode of the clade. Furthermore, within the Unionoidea, Cardioidea, Turbinoidea, Cerithioidea, Pleuroceridae, Littorinoidea, Muricoidea, Conoidea, and Succineidae, there are species with chromosome counts approximately half of the mode. Most of these clades are nested in larger clades with modes similar to their own.

Chromosome number has very high phylogenetic signal among the molluscan taxa studied, as demonstrated by two

Table 1
Maximum Likelihood Parameter Values for All Three Models and Sets of Branch Lengths

| Branch Lengths | Model | AIC | Root Count | $\lambda + \mu$ | $\lambda - \mu$ | δ |
|----------------------|-------------|---------|------------|-----------------|-----------------|----------------|
| Tree-10 ⁶ | Equilibrium | 364.540 | 17 | 5.0359 | 0 ^a | 0 ^a |
| | Trend | 366.504 | 16 | 4.9772 | 0.06851 | 0 ^a |
| | Duplication | 357.508 | 15 | 3.0555 | 0 ^a | 0.12304 |
| Tree-10 ⁵ | Equilibrium | 366.792 | 17 | 5.2310 | 0 ^a | 0 ^a |
| | Trend | 368.719 | 20 | 5.4529 | -0.34818 | 0 ^a |
| | Duplication | 361.744 | 15 | 3.4038 | 0 ^a | 0.11661 |
| Tree-10 ⁴ | Equilibrium | 369.134 | 17 | 5.2577 | 0 ^a | 0 ^a |
| | Trend | 371.043 | 20 | 5.4787 | -0.34931 | 0 ^a |
| | Duplication | 364.529 | 15 | 3.6157 | 0 ^a | 0.10860 |

NOTE.—All rates are in units of events/chromosome/billion years.

^a Rate is assumption of model, not set by maximum likelihood.

different parsimony statistics we derived from our data set and compared with the same statistics calculated for data sets generated by randomizing our data over the tips of the trees. Our data set had 250 steps for ordered parsimony and 13.36 steps for our weighted step matrix. Both these values were highly significant ($P \ll 0.0001$), as the 9,999 randomized data sets had from 558 to 749 steps for ordered parsimony and from 27.49 to 37.12 steps for our weighted step matrix.

Model Choice

The duplication model achieved its maximum likelihood at 15 ancestral chromosomes for all three trees, whereas the static model achieved its maximum likelihood at 17 (table 1). For both these models, likelihoods fell off precipitously for all trees when conditioned on ancestral chromosome counts that differed from their maximum likelihood chromosome count by more than 2 (fig. 3). The trend model reached its maximum likelihood at chromosome counts between 16 and 20 for the three different trees and although the likelihoods decreased precipitously for lower ancestral chromo-

some counts, they decreased more gradually for higher counts.

The AIC was much lower for the duplication model than it was for either other model when conditioned on ancestral chromosome counts near the maximum for all three trees (fig. 3). Furthermore, when the ancestral chromosome count was treated as an additional maximum likelihood parameter the AIC was much lower for the duplication model than it was for either other model (table 1). However, the trend model achieved relatively high maximum likelihoods over a much larger range of ancestral chromosome counts and when conditioned on more extreme chromosome counts it had lower AICs than either of the other two models (fig. 3). Nevertheless, the duplication model is clearly a better fit for the data as its AICs were so much lower for all three trees when the ancestral chromosome counts were fit by maximum likelihood.

There was little variation in the maximum likelihood estimates of parameter values between the three sets of branch lengths (table 1). Maximum likelihood estimates for the total rate of change for the birth–death process were

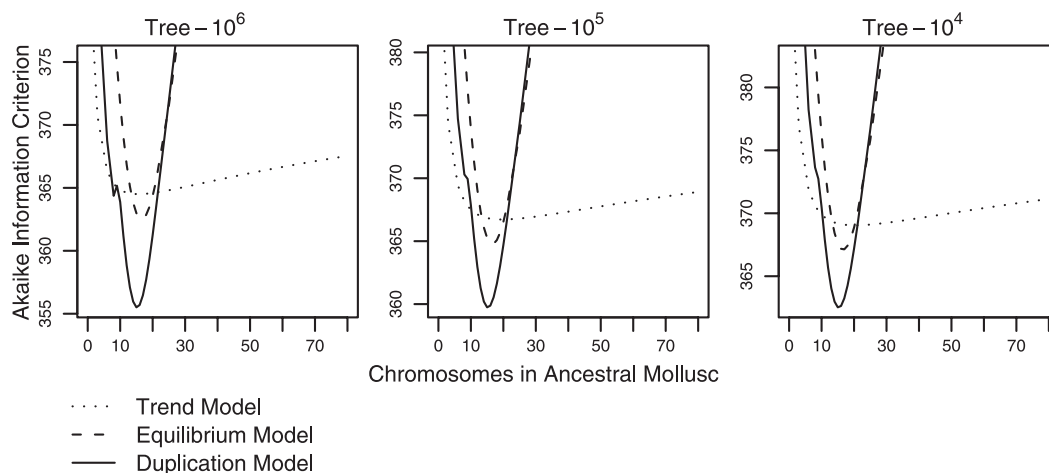


FIG. 3.—The AIC for each of three models of chromosome number evolution in extant mollusc families. Each figure shows the AICs for all three models conditioned on the number of chromosomes found in the last common ancestor of all Mollusca for a given set of branch lengths.

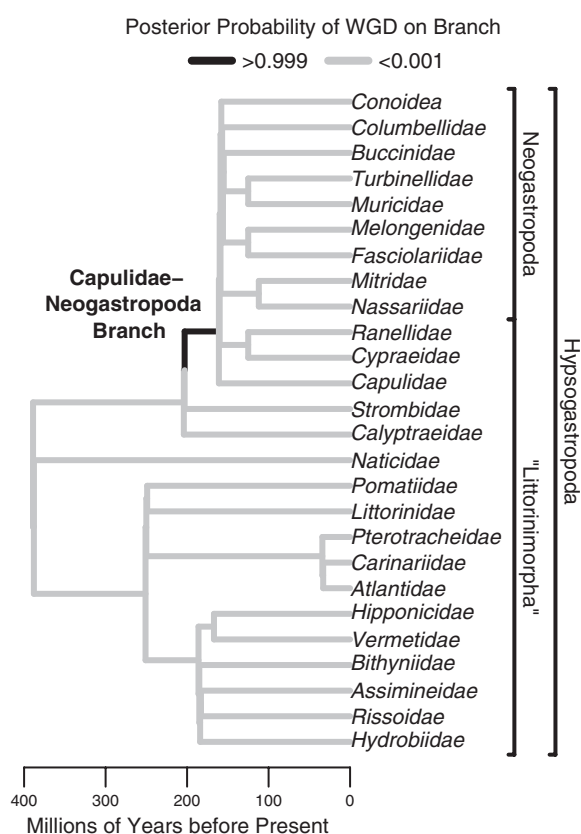


FIG. 4.—The phylogeny of all the terminal taxa in the Hypsogastropoda using the branch lengths from Tree-10⁶ showing the posterior probability of a WGD on each branch. The Capulidae-Neogastropoda branch is marked, as it has a very high posterior probability of a WGD for all sets of branch lengths, whereas all the other branches have essentially none.

much smaller under the duplication model than under the other two models because under the duplication model a great deal of the change in chromosome numbers can be accounted for by WGDs. The estimate of net change in chromosome number under the trend model was slightly positive for tree-10⁶; the estimate of net change was negative for the other two trees but still only represented 6.4% of the total rate. Our estimates of the WGD rate under the duplication model ranged from 0.109 duplications/billion years for Tree-10⁴ to 0.123 duplications/billion years for Tree-10⁶. It should be noted that this is the WGD rate that would be expected per lineage not over the whole tree; thus, we observed several WGDs in a clade with a 530 million year history.

Branches with Duplications

Three branches had posterior probabilities of WGDs greater than 0.67 for at least one set of branch lengths when conditioned on their maximum likelihood parameter values. These included the branch at the base of the coleoid cephalopods, a branch within the stylommatophoran gastropods at the base of a clade containing the Sigmurethra and the

Orthurethra and a branch within the hypsogastropods at the base of an unnamed clade. Several branches phylogenetically close to some of these well-supported branches also showed some support for a WGD (figs. 4–6). No other branch had a posterior probability greater than 0.03 for any of the trees.

The branch within the Hypsogastropoda at the base of a clade sister to the Strombidae and containing the Neogastropoda and several families of Littorinimorpha—hereafter to be referred to as the Capulidae-Neogastropoda branch—had posterior probabilities of a WGD greater than 0.999 for all three trees (fig. 4). This is extremely strong support for an evolutionary scenario in which the number of chromosomes doubled on this branch.

We calculated the posterior probability for a WGD on a branch within the Stylommatophora at the base of a clade containing the Sigmurethra and the Orthurethra of 0.881 for Tree-10⁶ (fig. 5). In contrast, Tree-10⁵ had much less support for a duplication on this branch and Tree-10⁴ had even less. Tree-10⁴ and Tree-10⁵ alternatively showed some support for a WGD on the branch at the base of the Stylommatophora, whereas Tree-10⁶ weakly supported a WGD on this branch. There was no support for WGDs occurring on both these branches; the posterior probability was less than 1×10^{-9} for all sets of branch lengths. The Sigmurethra-Orthurethra branch has a minimum branch length, and as a consequence, the likelihood of a duplication on this branch was less on trees with smaller minimum branch lengths. This greatly decreased the support for a doubling on this branch and instead compensated in part by increasing the support for a doubling on the ancestral stylommatophoran branch, which is 80 My long and thus much more likely to have a WGD on it. This appears to be an effect of our branch lengths, thus the analysis on Tree-10⁶, which is less affected by branch lengths, is likely more robust and suggests that a paleopolyploidy event occurred on the Sigmurethra-Orthurethra branch.

Our analysis clearly supports at least one paleopolyploidy event in the Cephalopoda, but it was difficult to distinguish whether there were one or two WGDs and the exact branch on which they occurred (fig. 6a). All three sets of branch lengths had a posterior probability of a WGD on the Coleoidea branch between 0.675 and 0.687 and on the Decapodiformes branch between 0.288 and 0.303 (fig. 6b), with weak support for duplications occurring on both of the branches (posterior probability between 0.030 and 0.046), meaning that the posterior probability that a WGD occurred on one of these branches was greater than 0.93 for all three sets of branch lengths. These trees also had posterior probabilities less than 0.09 that a WGD occurred either at the base of the Cephalopoda or on the Nautilidae branch with an overall posterior probability between 0.078 and 0.091 that a WGD occurred on more than one branch in this clade.

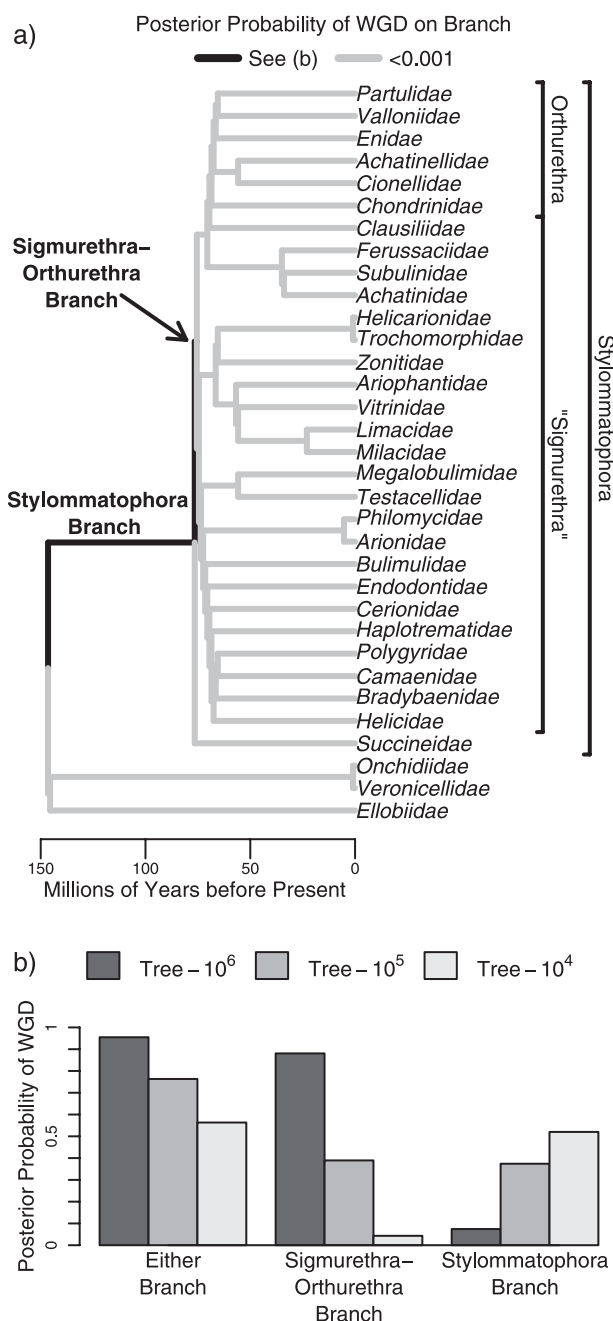


FIG. 5.—Posterior probabilities for a WGD on two branches near the base of the Stylommatophora. (a) The phylogeny of all the terminal taxa in a clade containing the Stylommatophora, the Systellommatophora, and the Elobiidae using the branch lengths from Tree-10⁶ showing the posterior probability of a WGD on each branch. The Stylommatophora branch and the Sigmurethra-Orthurethra branch are marked, as they have a relatively high posterior probability of a WGD, whereas all the other branches have essentially none. (b) A bar plot showing the posterior probability of a WGD on either of these branches as well as on each of these branches individually under each of the three different sets of branch lengths.

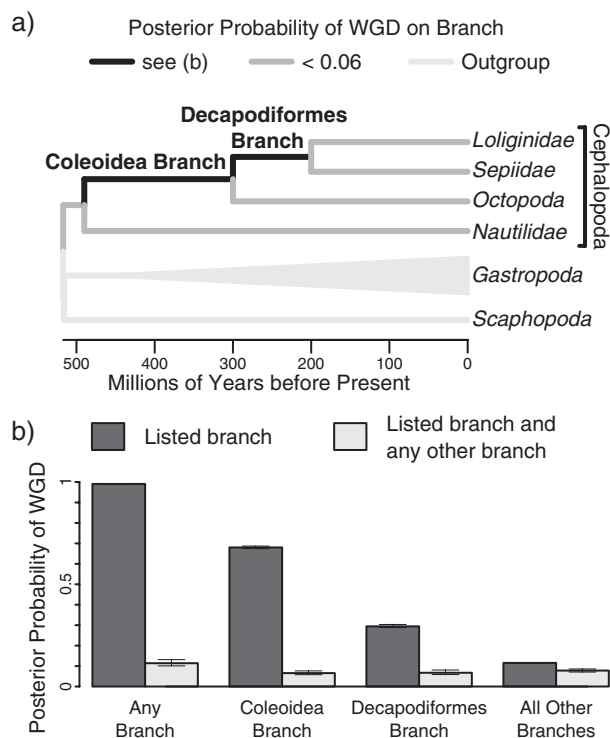


FIG. 6.—Posterior probabilities for a WGD on branches within the Cephalopoda. (a) The phylogeny of all the terminal taxa in the Cephalopoda with Gastropoda and Scaphopoda included as an outgroup using the branch lengths from Tree-10⁶ showing the posterior probability of a WGD on each branch. Two branches with relatively high posterior probability of a WGD are labeled above the branch. (b) A bar plot showing the posterior probability of a WGD on various branches in the phylogeny. The dark bars show the posterior probability of a WGD on the specified branch or branches. The light bars show the posterior probability that there were WGDs on any pair of branches in the Cephalopoda including the specified branch or branches. The posterior probability is shown by the large bar for Tree-10⁶, by the lower error bar for Tree-10⁵, by the upper error bar for Tree-10⁴.

Simulations to Evaluate Model Fit

We ran 1,000 simulations on Tree-10⁶ under the static model using our maximum likelihood parameter values for a process conditioned on 17 ancestral mollusc chromosomes. When evaluated under the static model, 160 of these simulations had maximum likelihoods lower than our data set, implying that we cannot reject this model as an explanation for our data ($P = 0.160$). We also calculated the maximum likelihood for all simulated data sets under the duplication model and used that value to calculate a log likelihood ratio. The largest log likelihood ratio for our simulated data sets was 5.506 and the likelihood ratio for over 96% of our simulations was less than 0.001, whereas the likelihood ratio for our actual data was 18.064. Thus, we can strongly reject our null hypothesis in favor of our alternative hypothesis of genome duplications ($P \ll 0.001$).

We also ran 1,000 simulations on Tree-10⁶ under the duplication model using the maximum likelihood parameter values for a process conditioned on 15 ancestral mollusc chromosomes. When evaluated under the duplication model, 40 of these simulations had maximum likelihoods less than our actual data, indicating that we can tentatively reject our alternative hypothesis as a fit for our model ($P = 0.040$).

However, one of the reconstructed genome duplications from our actual data appears to have occurred on a minimum length branch. Given the estimated value for δ of 0.123 WGD/billion years, the probability of a WGD on so short a branch is approximately 1.23×10^{-4} . Thus, any data set with a duplication on a minimum length branch will obviously have a lower maximum likelihood than one without. Given this estimated value of δ , we would expect WGDs to occur on a minimum length branch in 0.943% of our simulations, and indeed, they actually did occur in only nine of our simulations. Because the vast majority of our simulations did not have a duplication on a minimum length branch, we would expect them to have higher likelihoods than our actual data. Indeed, 166 of 1,000 simulations in which at least one duplication occurred on a minimum length branch had a maximum likelihood lower than our data set. Furthermore, 251 of 1,000 simulations in which duplications always occurred on the Capulidae–Neogastropoda branch, the Coleoidea branch, and the Sigmurethra–Orthurethra branch, and nowhere else had maximum likelihoods less than our actual data set. Thus, we could not reject the birth–death process as a model for our background rate of chromosome evolution.

Discussion

The likelihood model developed here and based on the birth–death process allows us to predict the phylogenetic position of paleopolyploidy events through comparative analysis of chromosome counts in extant species. When applied to a data set of chromosome numbers for the Mollusca, we found that a scenario in which the total number of chromosomes occasionally doubled explained these data better than a model in which chromosome number only evolved via the birth–death process (fig. 3). Furthermore, simulations indicated that the birth–death process was a poor fit for our chromosome number data when used alone, but a reasonable fit for the background rate of chromosome number evolution in the presence of WGDs. We identified three potential instances of paleopolyploidy (figs. 4–6). In one case, we could clearly identify the branch on which the WGD occurred; in the other two cases, we could narrow down the position of the WGD to one of two branches. Based on the assumptions inherent to our model, support for WGDs within the Mollusca in general and in these three clades in particular is substantial.

In the Caenogastropoda, comparative analysis of chromosome counts suggests that a WGD occurred in the common ancestor of a clade containing the Capulidae, the Ranellidae, the Cypraeidae, and the Neogastropoda after their divergence from the Strombidae and the other hypogastropod families included in our analysis (fig. 4). This paleopolyploidy event was strongly supported by all three sets of branch lengths (posterior probability > 0.999), and there was no support for a WGD on any other branch in the Hypogastropoda. Our interpretation of the fossil record indicates that this WGD occurred at some point between the beginning of the Jurassic (203 Ma) when the first Strombidae fossils appear and the lower Cretaceous (155 Ma) when the Neogastropoda initially radiated. Our results essentially agree with Hinegardner's (1974) suggestion that the neogastropods arose by polyploidy based on their genome size and chromosome number, although we inferred that this WGD occurred slightly earlier. If a WGD did occur at the base of the Neogastropoda, our phylogeny would imply two additional independent WGDs at the base of Capulidae and the Cypraeidae–Ranelidae. Considering the overall low rate of genome duplication in our analysis, such a scenario is highly unlikely. However, chromosome counts from groups not represented in this analysis (e.g., Ficidae and Cassidae) may affect this result.

We identified another likely WGD early in the history of the Stylommatophora: either at the beginning of the Cenozoic (65 Ma) in the common ancestor of the Sigmurethra and the Orthurethra after they diverged from the Succineidae or in the common ancestor of all the Stylommatophora after they diverged from the other pulmonates in the lower Cretaceous (138 Ma) and before they radiated at the beginning of the Cenozoic (65 Ma) (fig. 5a). Steusloff (1942) had also suggested a paleopolyploidy in the Stylommatophora, although we were able to more accurately locate the event. We could not establish a length for the Sigmurethra–Orthurethra branch or many of the branches immediately descended from it, as several extant Stylommatophoran families appear at the beginning of the Cenozoic and the phylogenetic position of Stylommatophora fossils appearing before then is uncertain (Zilch 1959–1960; Solem and Yochelson 1979) (fig. 2). Therefore, we treated each of these branches as one branch of minimum length within a relatively fast radiation. As a consequence, support for a paleopolyploidy event on the Sigmurethra–Orthurethra branch decreased for shorter minimum branch lengths, as the posterior probability of a WGD is proportional to the length of the branch. There was a concomitant increase in the posterior probability on the Stylommatophora branch, but it was not sufficient to compensate for the decrease on the Sigmurethra–Orthurethra branch, thus total support for a WGD on either of these branches decreased with shorter minimum branch lengths (fig. 5b). As the length of the Sigmurethra–

Orthurethra branch affects the posterior probability of both branches and the maximum support on the Sigmurethra–Orthurethra branch is higher than the maximum support for the Stylommatophora branch, we conclude that the lack of support for WGDs on the Sigmurethra–Orthurethra branch is in fact an artifact of our method of assigning branch lengths, and that it is more reasonable to conclude that a WGD occurred there than on the Stylommatophora branch. It should be noted that the number of chromosomes was highly variable among the Succineidae taxa (supplementary table, Supplementary Material online); thus, it is possible that selecting a different chromosome number to represent the Succineidae would lead to different conclusions about the location of the paleopolyploidy event. A well-resolved phylogeny of the Succineidae is necessary to better reconstruct the number of chromosomes in their last common ancestor but such a phylogeny is not currently available.

Our data suggested that a third WGD occurred within the Cephalopoda, although the exact location of this paleopolyploidy event is not clearly reconstructed (fig. 6a). All sets of branch lengths support a duplication in the common ancestor of the Coleoidea after they diverged from the Nautiloidea in the lower Ordovician (490 Ma) but before the Decapodiformes split from the Octopoda in the Carboniferous (300 Ma). On the other hand, there is also some support for a duplication occurring in the common ancestor of the Decapodiformes after they split from the Octopoda but before the Sepiidae and the Loliginidae split in the lower Jurassic (200 Ma). There is also weak support for multiple WGDs within the Cephalopoda either on both these branches or on some combination of these branches and other branches within the Cephalopoda (fig. 6b). Inspection of chromosome counts within the Cephalopoda implies another scenario. The Octopoda have 30 chromosomes in their haploid genome on the other hand the Sepiidae and the Loliginidae both have 46. Given that we reconstructed 15 chromosomes in the ancestral mollusc, this implies that the Octopoda are tetraploid, whereas the Decapodiformes are hexaploid. Our model does not account for changes in ploidy other than doublings and thus would be limited to identifying tetraploids and octoploids. The hexaploid Decapodiformes likely caused the uncertainty in reconstructing the WGD within the Cephalopoda. Although variation in chromosome counts within both the Sepiidae and the Loliginidae is overall very low, there are taxa in both families that have been reported to have many more chromosomes than the mode for the family (supplementary table, Supplementary Material online), such as *Loligo plei* (84) and *Sepia officinalis* (56). If these values are correct it is possible that the ancestral Decapodiformes had more chromosomes than are suggested by our reconstructions. This would increase support for a WGD on the Decapodiformes branch.

It has been suggested that WGDs can lead to large morphological and physiological innovations, as redundancy eases the constraints on genes throughout the genome (Haldane 1932; Ohno 1967). The three clades recognized here fit this proposition well. The Stylommatophora represent the dominant group of land snails and slugs and have successfully made the difficult transition from an aquatic to terrestrial life style and all the associated anatomical and physiological changes that this transition requires (Barker 2001; Mordan and Wade 2008). The coleoid cephalopods are also a large group with a diversity of body forms and ecologies and represent a large jump in anatomical, physiological, and behavioral complexity. They are characterized by an internalized or reduced shell, a muscular mantle for locomotion, chromatophores, ink sacs, an acute vision system, and complex and morphologically distinct arms (Nishiguchi and Mapes 2008). Lastly, according to our reconstructions, the WGD event identified in the Caenogastropoda precedes the radiation of the Neogastropoda by only a short time and may have played a major role in their diversification. Neogastropoda are a large and diverse clade that have undergone extraordinary radiations as seen in their anatomical, physiological, behavioral, and ecological diversity (Ponder et al. 2008).

Although the WGDs suggested here are the only well-supported paleopolyploidies in the Mollusca that we are aware of, more recent polyploidies have long been recognized in many species of molluscs. Several polyploid species have been identified in the wild (e.g., Patterson and Burch 1978; Barsiene et al. 1996; Park 2008) and in at least one case the origin of a polyploidy has been analyzed karyologically (Goldman et al. 1983). Polyploidy has also been artificially induced in several species of commercial molluscs (e.g., Beaumont and Fairbrother 1991; Yang and Guo 2006; Le Pennec et al. 2007). Furthermore, polyploidy is common in the somatic tissue of many Molluscs (Anisimov et al. 1995; Tabakova et al. 2005; Tokmakova et al. 2006). Within several of the terminal taxa in our analysis, we recognized members with approximately twice the number of chromosomes as the mode for that family, and these may also be a consequence of more recent WGDs (supplementary table, Supplementary Material online).

We observed a strong phylogenetic signal in mollusc chromosome number with no tendency for the number of chromosomes to increase or decrease. Some researchers had previously noticed that taxonomic groups within the Mollusca tended to have similar numbers of chromosomes (Patterson and Burch 1978; Nakamura 1985), as our analysis confirmed ($P \ll 0.001$). However, others have observed a tendency for chromosome numbers to either increase (e.g., Patterson and Burch 1978) or decrease (e.g., Butot and Kiauta 1969; Ahmed 1976). The maximum likelihood analysis presented here suggests that there is no trend in chromosome number, although the WGDs could be

construed as representing a positive trend. Vinogradov (2000) detected a general tendency to large genome sizes in terrestrial pulmonates. However, as his analysis was not phylogenetic, the signal may be a consequence of large chromosome counts in the Stylommatophora.

Mayrose et al. (2010) described a similar method to deduce the phylogenetic location of paleopolyploidies and used a maximum likelihood implementation of it to analyze chromosome counts from several plant taxa. The theory behind both methods is essentially the same, although they established a rate matrix and approximated the exponent of that rate matrix to calculate the transition probabilities. Thus, they were able to easily use a wider variety of models including models similar to those used here and models in which the rates of chromosome increase and decrease were independent of the number of chromosomes, which was a slightly better fit for their data. They were also able to include rates for genome doubling, genome halving, and genome increasing by 50% and to exclude the possibility of reaching zero chromosomes from their models. Our research was almost complete at the time that Mayrose et al. (2010) was published, and so our analysis does not incorporate their methods, but the two methods are very similar and likely to yield the same results. Here, we established an elaborate hypothesis testing framework using simulations to confirm the fit of our model to the data set, which could easily be incorporated into the method described by Mayrose et al. (2010).

The methodology we have presented here is likely applicable to other taxa provided that the background rate of chromosome number evolution is low enough relative to the time spans involved. However, background rates of change in chromosome number in the Mollusca and, particularly, in gastropods may be unusually low (Chambers 1987). In other taxa, the background rate may drown out the signal of WGDs. It would also be possible to apply this method to counts of gene family members or large syntenic regions, which would also be expected to double after a WGD. Much support for paleopolyploidy has come from the identification of repeated syntenic regions (e.g., Holland et al. 1994; Abi-Rached et al. 2002; Wong et al. 2002).

The results presented here represent only the first step in the study of molluscan paleopolyploidies, and we hope it may help guide future research into molluscan polyploidy and in the selection of molluscan taxa for whole genome sequencing (fig. 1). Ideally, the study of WGDs requires multiple genome sequences (Wong et al. 2002; Woods et al. 2005; Scannell et al. 2007), but currently there exist less than five genome sequencing and annotation efforts within the Mollusca (Chapman et al. 2007). As costs decrease and bioinformatic methodologies continue to rapidly advance, the method we outline here may prove useful in selecting taxa for the study of molluscan paleopolyploidy. In addition, researchers interested in paleopolyploidy in non-molluscan

taxa may also find the methods developed here useful in their selection of taxa for whole genome sequencing by identifying branches on which the number of chromosomes or members of gene families have doubled.

Supplementary Material

Supplementary table is available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We would like to thank the University of California, Berkeley, for funding the open access costs of this research through the Berkeley Research Impact Initiative.

Literature Cited

- Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat Genet.* 31(1):100–105.
- Ahmed M. 1976. Chromosome cytology of marine pelecypod molluscs. *J Sci (Karachi).* 4:77–94.
- Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 19(6):716–723.
- Anisimov AP, Tokmakova NP, Poveschenko OS. 1995. Somatic polyploidy in some tissues of the succineid snail *Succinea lauta* (Gastropoda: Pulmonata). *Tsitologiya* 37(4):311–330.
- Ansorge WJ. 2009. Next-generation DNA sequencing techniques. *N Biotechnol.* 25(4):195–203.
- Bailey NTJ. 1964. The elements of stochastic processes. New York: John Wiley & Sons.
- Barker GM. 2001. Gastropods on land: phylogeny, diversity and adaptive morphology. In: Barker GM, editor. *Biology of terrestrial molluscs*. Wallingford (CT): CAB International. p. 1–146.
- Barsiene J, Tapia G, Barsyte D. 1996. Chromosomes of molluscs inhabiting some mountain springs of eastern Spain. *J Molluscan Stud.* 62(4):539–543.
- Beaumont AR, Fairbrother JE. 1991. Ploidy manipulation in molluscan shellfish: a review. *J Shellfish Res.* 10:1–18.
- Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16(7):1667–1678.
- Butot LJM, Kiauta B. 1969. Cytotaxonomic observation in the stylommatophoran family Helicidae, with considerations on the affinities within the family. *Malacologia* 9:261–262.
- Byrne KP, Blanc G. 2006. Computational analyses of ancient polyploidy. *Curr Bioinform.* 1(2):131–146.
- Chambers S. 1987. Rates of evolution in chromosome numbers in snails and vertebrates. *Evolution* 41(1):166–175.
- Chapman JA, Edsinger-Gonzales E, Begovic E, Lindberg DR, Rokhsar DS. 2007. The genome of the owl limpet: a model for molluscs? *International Plant & Animal Genomes XV Conference*; 2007 Jan 13–17; San Diego, CA. San Diego (CA).
- Colgan DJ, Ponder WF, Beacham E, Macaranas J. 2007. Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca). *Mol Phylogenet Evol.* 42(3):717–737.

- Felsenstein J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst Biol.* 22(3):240–249.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* 17(6):368–376.
- Footo M, Hunter JP, Janis CM, Sepkoski JJ Jr. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* 283(5406):1310–1314.
- Gao Y, Natsukari Y. 1990. Karyological studies on seven cephalopods. *Venus* 49(2):126–145.
- Giribet G, Wheeler W. 2002. On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. *Invertebr Biol.* 121(4):271–324.
- Goldman MA, Loverde PT, Chrisman CL. 1983. Hybrid origin of polyploidy in fresh water snails of the genus *Bulinus* Mollusca Planorbidae. *Evolution* 37(3):592–600.
- Hahn M, De Bie T, Stajich J, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* 15(8):1153–1160.
- Haldane JBS. 1932. The causes of evolution. Ithaca (NY): Cornell University Press.
- Hallinan N. 2011. Using the birth-death process to infer changes in the pattern of lineage gain and loss. [dissertation]. [Berkeley (CA)]: University of California-Berkeley.
- Haszprunar G. 2000. Is the Aplousobranchia monophyletic? A cladistic point of view. *Am Malacol Bull.* 15:115–130.
- Hinegardner R. 1974. Cellular DNA content of the Mollusca. *Comp Biochem Physiol A Physiol.* 47(2):447–460.
- Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. *Dev Suppl.* 1994:125–133.
- Ieyama H. 1993. Karyotypes of two species in the dentaliidae (Mollusca: Scaphopoda). *Venus* 52(3):245–248.
- Kasahara M. 2007. The 2R hypothesis: an update. *Curr Opin Immunol.* 19(5):547–552.
- Kendall DG. 1948. On the generalized “birth-and-death” process. *Ann Math Statist.* 19(1):1–15.
- Le Pennec G, et al. 2007. Polyploidisation and gametogenesis in a cultivated bivalve mollusc, Ostreid *Crassostrea gigas*. *Boll Malacol.* 43(1–8):87–95.
- Lindgren AR, Giribet G, Nishiguchi MK. 2004. A combined approach to the phylogeny of the Cephalopoda (Mollusca). *Cladistics* 20(5):454–486.
- Lynch M, Conery J. 2003. The evolutionary demography of duplicate genes. *J Struct Funct Genomics.* 3(1):35–44.
- Mayrose I, Barker M, Otto S. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst Biol.* 59(2):132–144.
- McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. *Nat Genet.* 31(2):200–204.
- Mishler B. 1994. Cladistic-analysis of molecular and morphological data. *Am J Phys Anthropol.* 94(1):143–156.
- Mordan P, Wade C. 2008. Heterobranchia II. The pulmonata. In: Ponder WF, Lindberg DR, editors. *Phylogeny and evolution of the mollusca*. Berkeley (CA): University of California Press. p. 409–426.
- Nakamura HK. 1985. A review of molluscan cytogenetic information based on the CISMOCH-Computerized Index System for molluscan chromosomes. Bivalvia, Polyplacophora and Cephalopoda. *Venus.* 44(3):193–225.
- Nakamura HK. 1986. Chromosomes of archaeogastropoda (Mollusca: Prosobranchia), with some remarks on their cytotaxonomy and phylogeny. *Publ Seto Mar Biol Lab.* 31:191–267.
- Nishiguchi MK, Mapes RH. 2008. Cephalopoda. In: Ponder WF, Lindberg DR, editors. *Phylogeny and evolution of the mollusca*. Berkeley (CA): University of California Press. p. 163–200.
- Nixon K, Crepet W, Stevenson D, Friis E. 1994. A reevaluation of seed plant phylogeny. *Ann Missouri Bot Gard.* 81:484–533.
- Novozhilov AS, Karev GP, Koonin EV. 2006. Biological applications of the theory of birth-and-death processes. *Brief Bioinform.* 7(1):70–85.
- Odierna G, et al. 2008. Karyology of the Antarctic chiton *Nuttallochiton mirandus* (Thiele, 1906)(Mollusca: Polyplacophora) with some considerations on chromosome evolution in chitons. *Chromosome Res.* 16(6):899–906.
- Ohno S. 1967. Sex chromosomes and sex-linked genes. Berlin (Germany): Springer-Verlag.
- Okusu A, Schwabe E, Eernisse DJ, Giribet G. 2003. Towards a phylogeny of chitons (Mollusca, Polyplacophora) based on combined analysis of five molecular loci. *Org Divers Evol.* 3(4):281–302.
- Park GM. 2008. Polyploidy in three sphaeriids (Bivalvia: Veneroidea) from Korea. *Molluscan Res.* 28(2):133–136.
- Patterson CM. 1969. Chromosomes of molluscs. In: *Proceedings of the 2nd Symposium of Mollusca*. Vol. 2. 1969; Ernakulam, Cochin (India). Colchin (India): Marine Biological Association of India. p. 635–689.
- Patterson CM, Burch JB. 1978. Chromosomes of pulmonate molluscs. In: Fretter V, Peake J, editors. *Pulmonates*. Vol. 2A, Systematics, evolution and ecology. New York: Academic Press. p. 171–217.
- Piskur J. 2001. Origin of the duplicated regions in the yeast genomes. *Trends Genet.* 17(6):302–303.
- Ponder W, et al. 2008. Caenogastropoda. In: Ponder WF, Lindberg DR, editors. *Phylogeny and evolution of the mollusca*. Berkeley (CA): University of California Press. p. 331–383.
- Ponder WF, Lindberg DR. 1997. Towards a phylogeny of gastropod mollusc: an analysis using morphological characters. *Zool J Linn Soc.* 119(2):83–265.
- Scannell D, et al. 2007. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc Natl Acad Sci U S A.* 104(20):8397–8402.
- Semon M, Wolfe KH. 2007. Consequences of genome duplication. *Curr Opin Genet Dev.* 17(6):505–512.
- Solem A, Yochelson EL. 1979. North American Paleozoic land snails, with a summary of other Paleozoic nonmarine snails. U S Geological Survey Professional Paper 1072. p. 1–42.
- Spring J. 1997. Vertebrate evolution by interspecific hybridisation—are we polyploid? *FEBS Lett.* 400(1):2–8.
- Steusloff U. 1942. Weitere Beiträge zur Kenntnis der Verbreitung und Lebensansprüche der *Vertigo genesii-parcedentata* im Diluvium und Alluvium. (Polyploidie whrend des Periglazias?). *Arch Molluskenk.* 74:192–212.
- Tabakova EV, Kirsanova IA, Anisimov AP. 2005. Morphological variability and ploidy of nuclei in neurons of the central nervous system of bivalves in connection with somatic polyploidy. *Russ J Mar Biol.* 31(5):352–357.
- Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet.* 38(1):615–643.
- Thiriot-Quievreux C. 2002. Review of the literature on bivalve cytogenetics in the last ten years. *Cah Biol Mar.* 43(1):17–26.
- Thiriot-Quievreux C. 2003. Advances in chromosomal studies of gastropod molluscs. *J Molluscan Stud.* 69(3):187–201.

- Tokmakova NP, Galimulina NA, Anisimov AP. 2006. Morphofunctional characterization and ploidy levels of cells of the bivalve digestive gland with special reference to somatic polyploidy. *Russ J Mar Biol.* 32(4):270–276.
- Vandepoole K, Vos WD, Taylor JS, Meyer A, de Peer YV. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc Natl Acad Sci U S A.* 101(6):1638–1643.
- Vinogradov AE. 2000. Larger genomes for molluscan land pioneers. *Genome* 43(1):211–212.
- Vitturi R, Colombera D, Catalano E, Arnold J. 1990. Spermatocyte chromosome study of eight species of the class Cephalopoda (Mollusca). *J Cephalopod Biol.* 1:101–112.
- Wade CM, Morton PB, Naggs F. 2006. Evolutionary relationships among the pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biol J Linn Soc Lond.* 87(4):593–610.
- Wägele H, Klussmann-Kolb A, Vonnemann V, Medina M. 2008. Heterobranchia I. The opisthobranchia. In: Ponder WF, Lindberg DR, editors. *Phylogeny and evolution of the mollusca*. Berkeley (CA): University of California Press. p. 385–408.
- Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol Biol.* 42(1):225–249.
- White MJD. 1973. *Animal cytology and evolution*. Cambridge: Cambridge University Press.
- Wolfe K. 2004. Evolutionary genomics: yeasts accelerate beyond BLAST. *Curr Biol.* 14(10):R392–R394.
- Wong S, Butler G, Wolfe KH. 2002. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc Natl Acad Sci U S A.* 99(14):9272–9277.
- Woods IG, et al. 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.* 15(9):1307–1314.
- Yang H, Guo X. 2006. Polyploid induction by heat shock-induced meiosis and mitosis inhibition in the dwarf surfclam, *Mulinia lateralis* Say. *Aquaculture* 252(2–4):171–182.
- Zilch A. 1959–1960. Euthyneura. In: Schindewolf O, editor. *Handbuch der Paläozoologie*, volume 6, chapter 2. Berlin (Germany): Gebrüder Bornträger.

Associate editor: Geoff McFadden