Analysis of Salamander Genome Size

Genome size is important in all animals.
Genome size is important in salamanders.
Salamanders provide a unique opportunity to test evolutionary hypotheses of genome size evolution.

Genome size is thought to be under selection because duplication of the genome is the rate-limiting step in cell division. Therefore animals that require fast cell division (i.e., rapid growth associated with high metabolic rate) should experience selective pressure for small genome size (Edwards and Waltari 2003, others?). Another factor thought to be involved is cell volume. Genome size is a primary determinant of nucleus size, which is strongly correlated with cell size. Because animals with high metabolic rate require large surface area to volume ratios of their cells to support their high metabolic rates, they should experience a downward selective pressure on genome size.

By similar logic, animals that have no particular need for rapid cell division may be released from selection and free to expand their genome sizes.... However, it is not known whether selection or constraints predominate the evolutionary dynamics of genome size evolution. Furthermore, the the extent that selection plays a role, it is unknown which aspects of life history provide strong selective pressures.

Salamanders have the largest range of genome size variation known in the animal kingdom (true?). In particular, they have some of the largest genome sizes known amongst vertebrates, orders of magnitude larger than even other ectotherms.

Ectotherms, and salamanders in particular, may be released from these selective pressures (or at least experience relaxed selection) because of their low-metabolic rate, slow-growth life histories. Salamanders generally go through a biphasic life cycle with aquatic larvae and terrestrial adults. The time at which they are they would experience the most pressure for rapid development is during the transition from larvae to adult. During this time, they do not feed and furthermore their locomotion may be suboptimal as they transition between environments and from a swimming to a walking morphology. It is reasonable to assume, therefore, that metamorphic salamanders may be under more pressure for smaller genome size than non-metamorphs. Conversely, non-metamorphic salamanders may be released from this constraint on genome size.

In addition to metamorphic salamanders, there are paedomorphs which retain the larval form through adulthood, and direct developers, which never have a larval phase. These life histories are expected to have less constraint on the upper limits of genome size. Each life history may have a distinct optimum genome size, or they may share a general optimum.

Finally, it has long been recognized that plethodon salamanders have some of the largest genome sizes in the animal kingdom. Most plethodons are direct developers, and therefore bypass the vulnerable metamorphic transition. Therefore there should be

no point in their life history that requires rapid development. Plethodons are unique in many mays, including being lungless, which implies very low metabolic rate. They also have reduced cell number(?) resulting in fewer, larger cells and therefore less constraint on genome size(?). However, within plethodons, metamorphosis has re-evolved from direct developing ancestors multiple times. This provides a unique opportunity to explore selective pressures from multiple major life-history transitions.

In this study, we explore the evolutionary influences on genome size using salamanders as a model system. We explore complex models of adaptive evolution with multiple optima representing alternative scenarios of the importance of selective factors using Orstein-Ulhenbeck models (OU), versus models of neutral evolution with multiple rates of stochastic evolution representing relaxation from constraint using the multiple-rate Brownian motion models (BM). In particular, we explore whether metamorphic or alternative life histories, the unique biology of plethodons, or their interaction, have shaped the evolution of genome size.

Methods

Genome Size and Taxa Included

We included 106 species of salamander in this study, representing metamorphs, paedomorphs, direct developers, plethodons, and metamorphosing plethodons. We obtained genome size data from XX, and is given in pg (table 1). Because of the very large range of variation in genome size, we transformed the data with natural logarithms prior to analysis. Natural log-transformation is a reasonable choice in this case, because genome size expansion is thought to be accelerated through molecular biology of repetitive elements, with larger repetitive elements giving rise to larger changes in genome size. [[need to describe the mechanism with a sentence]].

Taxa Included and Phylogeny

The phylogeny was obtained by....

Models of Evolution

Hypotheses of genome size evolution were tested using both Brownian motion (BM) and Ornstein-Uhlenbeck (OU) models of evolution. These models were discussed previously (Hansen 1997; Butler and King 2004; O'Meara et al. 2006; Beaulieu et al. 2012), and thus we provide only a brief description here. The standard BM model describes the evolution of a quantitative trait in terms of stochastic variation as a function of time. As a stochastic differential equation it can be expressed as:

$$dX(t) = sigma\ dB(t)\ dt,$$

where the phenotypic trait X(t) evolves through time t, with random deviations introduced by the white noise variable dB(t). The magnitude of stochastic deviation is influenced by the sigma parameter, which is often called the “rate of evolution” in the literature. The white noise variable *dB(t)* can be thought of as a draw from a random normal distribution with mean zero and variance *dt*.

The multiple-rate BM model was introduced by O’Meara et al (2006) and is slightly more complex than the standard BM model in that it allows different sigma parameters in different portions of the phylogenetic tree. This can be used, for example, to test whether there has been a relaxation ...

Evolutionary Hypotheses for the Evolution of Genome Size

Evolutionary hypotheses for the evolution of genome size fell into two major categories: those that relate to alternative life histories in salamanders (metamorphosis, direct development, paedomorphosis; Figure 1A), and those that relate to the plethodon clade (Figure 1B). Plethodons lost metamorphsis early in their evolutionary history, with most species adopting a direct-developing strategy. In fact, all direct developing salamanders are plethodons. Those plethodons that are metamorphic have therefore regained metamorphosis. Metamorphosis in plethodons, however, is distinct from the ancestral form of metamorphosis. It has been characterized as “radical” -- plethodons cannot feed during this time, and therefore metamorphose quickly relative to other salamanders.

1) *metamorphosis*: This hypothesis tests whether metamorphosis, in general, imparts selection for smaller genome size.
2) *mpd*: Metamorphosis vs. paedomorphosis vs. direct development. The distinct life histories result in distinct evolutionary optima for genome size. This hypothesis tests whether the loss of metamorphosis is the key variable and whether the way in which metamorphosis is lost (i.e. paedomorphosis, direct development) matters for optimal genome size.
3) *mpdMpleth*: Same as for mpd, with the addition that the regain of metamorphosis within plethodons imparts a distinct selective regime.
4) *MxMpleth*: Metamorphosis, and metamorphosis within plethodons are distinct selective regimes separate from non-metamorphosers.
5) *xMpleth*: This tests whether metamorphosis in plethodontids, specifically, imparts selection for smaller genome size. This makes biological sense because plethodontid metamorphosis is radical.
6) *xpleth*: This hypothesis implies that there is some aspect of plethodon life history that imparts a distinct selective pressure, regardless of life history mode.
7) *xplethMpleth*: Non-plethodontids v. non-metamorphosing plethodontids vs. metamorphosing plethodontids. This tests whether the evolution of "radical" metamorphosis at the base of the plethodontid clade had a lasting impact on genome size evolution in the plethodontid clade.

8) *xplethMplethPleth*: Same as previous, but with an additional regime for paedomorphic plethdons.

9) *global*: A single, shared selective regime for all salamanders.

10) *BM*: Brownian motion model with no selective optima. Evolution in genome size is dominated by stochastic mechanisms, or possibly adaptive factors not modeled above.

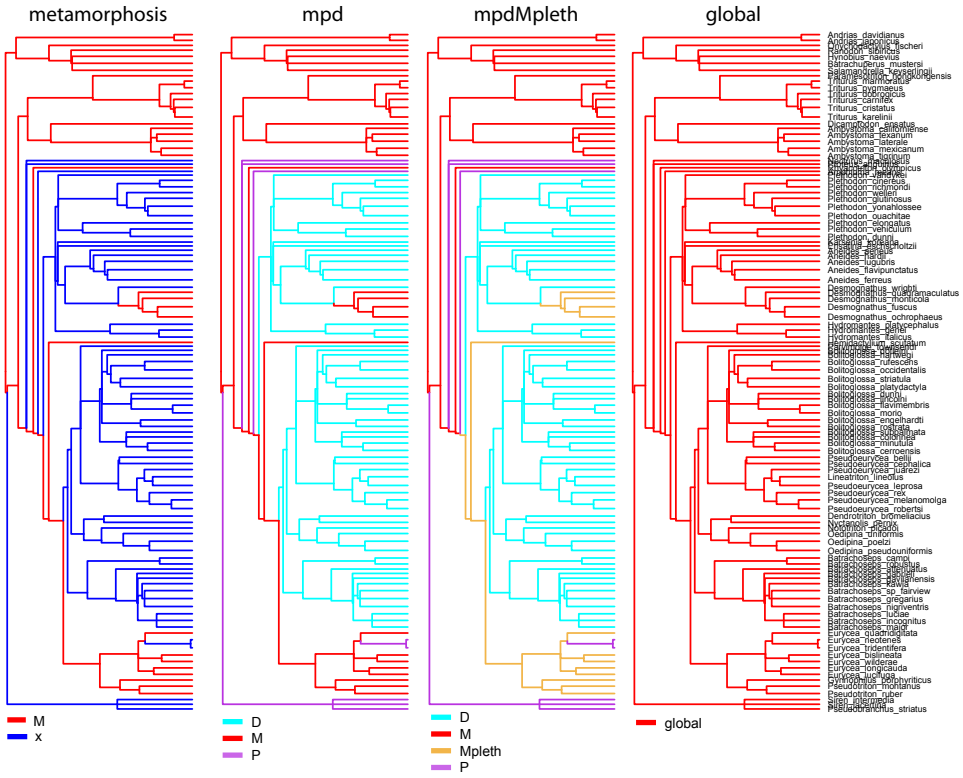

Figure 1A. Evolutionary hypotheses for genome size.

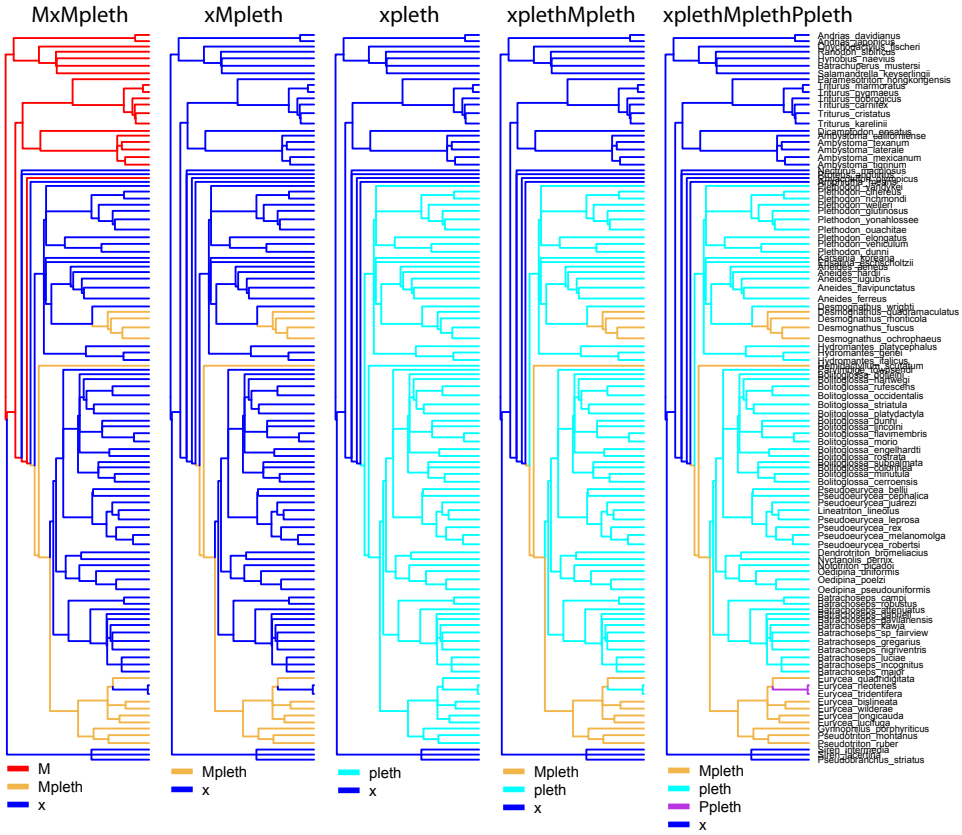

Figure 1B. More evolutionary hypotheses for genome size.

Analyses

The evolutionary hypotheses were tested both as alternative sets of evolutionary optima using the multiple-optimum OU model, or as shifts in rate in the mutiple-rate BM model. Additional alternatives included the single-rate BM model and the single optimum OU model (global optimum model). The BM and multiple-optimum OU models were implemented using the software package ouch (Butler and King 2004; King and Butler 2009). Single and multiple-rate BM models were tested using the OUwie package (O'Meara et al 2006; Beaulieu and O'Meara 2013). Both packages were written in the R statistical computing environment (R Core Team 2013).

Is there anything interesting to say about paedomorphic plethodons? If not, I think we should drop the xplethMplethPpleth model -- it only has 2 species??

Results

The best model overall is the mpdMpleth multiple optimum model (metamorphosis, paedomorphosis, direct development, and metamorphic plethodons; see Table 1 or alternative Table 1). The best model was supported both by the lowest AIC value, as well as the bootstrap model selection statistics. The mpdMpleth model was selected 61% of the time when competed against the remaining multiple optimum models in parametric bootstrap simulations. The second best model was selected only 10% of the time.

The mpdMpleth model was much better than the mpd model, which indicates the the regain of metamorphosis in plethdons results in a distinct selective optimum for genome size than that of metamorphosis in the ancestral condition. It is also interesting that the metamorphosis model is much worse than the mpd model, which shows that distinguishing paedomorphs from direct developers is needed to improve fit.

Based on the best model, one may suppose that including an Mpleth optimum is really important. However, the model xMpleth is worse, which splits metamorphic plethodons from all other taxa, and is similar to a BM model. The explanatory power also does not simply stem from being a plethodon either, since the xpleth model fits poorly. The xplethMpleth model is as good as xMpleth and BM, and splitting out the Ppleths makes the model fit worse, so the Ppleths provide no additional explanatory power. Metamorphosis, metamorphic plethodons, with all other taxa grouped together (MxMpleth) is also worse.

Table 1. Model comparison statistics for multiple optimum OU models.

	aic	aic.c	sic	dof
mpdMpleth	10.19	11.03	26.17	6
BM	12.62	12.73	17.95	2
xMpleth	12.81	13.21	23.47	4
xplethMpleth	12.81	13.41	26.13	5
mpd	12.89	13.49	26.21	5
MxMpleth	13.10	13.70	26.42	5
metamorphosis	13.94	14.34	24.59	4
xplethMplethPpleth	14.75	15.60	30.73	6
xpleth	17.62	18.01	28.27	4
OU1	20.09	20.33	28.08	3

Alternative

NOTE: I played around with the categories some more, and if we take mpd, but then make a category called ND (pooling the Mpleth and Ppleth), we get a better fit. The AIC drops to 9.55.

Table 1. Model comparison statistics for multiple optimum OU models and multiple rate BM models. Model selection bootstrap for multiple optimum OU models given in parentheses. Percentages indicate the frequency of selecting each as the best model.

	AIC	
	OU	BM
	Multiple Optimum	Multiple Rate
mpdMpleth	**10.19 (61%)**	17.10
BM	12.62 (6%)	12.62
xMpleth	12.81 (6%)	13.17
xplethMpleth	12.81 (6%)	15.04
mpd	12.89 (10%)	16.51
MxMpleth	13.10 (4%)	15.12
metamorphosis	13.94 (3%)	14.50
xplethMplethPpleth	14.75 (3%)	16.64
xpleth	17.62 (1%)	14.41
OU1	20.09 (0%)	20.09

Interestingly, BM is not as terrible as some of the other adaptive models. In particular the single-optimum OU (OU1) was the worst-fitting. Thus, there is clear support for more than one optimum, and furthermore, it is important to include an adaptive component, even if the improvement is modest. However, the adaptive component has to identify the correct selective regimes in order to be informative.

Based on the rank of the BM model amongst other models, a reasonable question is whether a multiple rate BM model would provide more explanatory power. However, all of the multiple-rate BM models provided worse fits than their multiple optimum counterparts (Alt. Table 1). The single exception is the xpleth model, however both versions of this model are poorly fitting.

The parameters estimated from the best fit model indicate a model with weak selection and weak drift (Table 2). The estimated evolutionary optima are reasonable for the direct developers and the metamorphosers, however, they are very far from reasonable for the remaining groups, with an exceedingly large value for paedomorphs and a very small value for metamorphic plethodons. We have no obvious explanation for this, but it does suggest some strain. The fact that the parameter estimates for the optima have such very large 95% confidence intervals shows that these parameters are difficult to estimate, and that we should not put much confidence in these point estimates. Although the lower bounds vary considerably, there is much overlap in the confidence intervals.

Table 2. Parameter estimates for best model, the multiple optimum mpdMpleth. The optima are back-transformed to pico grams of DNA.

	Parameter Estimate	95% CI
sigma.squared	0.12	(0.10, 0.25)
alpha	0.09	(0.15, 1.87)
optima D	30.19	(4.37, 231)
optima M	37.40	(8.22, 164)
optima P	8077	(19.23, 2436)
optima Mpleth	0.05	(0.27, 67)

It is odd that the parameter estimates are outside of the 95% CI's for P and Mpleth. I don't know what to make of that either.

I still don't know what to do about the SIC discrepancy. Maybe ignore it?

4) When we look at SIC, things flip around. It is known that SIC has a higher penalty for additional parameters than AIC, and the mpdMpleth model has 4 more parameters than BM. I don't know what to make of that exactly, but when I look at the parameters, we are in a region with weak selection. So maybe the signal is strong enough to pick up with AIC but not with SIC. I didn't do all the bootstraps yet but I remember that the AIC results are robust to parametric boot strap. So we just don't get the same answer with AIC and SIC. But that's going to have to be OK. We should think if this is consistent with weak selection. I think so.