1-16-2013

Dictyostelium development shows a novel pattern of evolutionary conservation

Xiangjun Tian
Washington University in St Louis

Joan E. Strassmann
Washington University in St Louis, strassmann@WUSTL.EDU

David C. Queller
Washington University in St Louis, queller@WUSTL.EDU

Dictyostelium development shows a novel pattern of evolutionary conservation

Journal:	*Molecular Biology and Evolution*
Manuscript ID:	MBE-12-0588.R2
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Tian, Xiangjun; Washington University in St. Louis, Biology Department CB1137 Strassmann, Joan; Washington University in St. Louis, Biology Department CB1137 Queller, David; Washington University in St. Louis, Biology Department CB1137
Key Words:	evolution, development, constraint, evolvability, Dictyostelium, hourglass model

Dictyostelium development shows a novel pattern of evolutionary conservation

The submission is intended as an Article

Xiangjun Tian, Joan E. Strassmann, and David C. Queller[1]

Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA

[1]Corresponding author:

David C. Queller
Department of Biology CB 1137
Washington University in St. Louis
One Brookings Drive
St. Louis MO 63130 U.S.A.
Tel: (+1) 314-935-3528
Fax: (+1) 314-935-4432
Email: queller@wustl.edu

Key words: von Baer's law, hourglass model, *Dictyostelium*, evolutionary conservation, development

Running head: Evolutionary conservation of *Dictyostelium* development

Abstract

Von Baer's law states that early stages of animal development are the most conserved. More recent evidence supports a modified "hourglass" pattern in which an early but somewhat later stage is most conserved. Both patterns have been explained by the relative complexity of either temporal or spatial interactions; the greatest conservation and lowest evolvability occur at the time of the most complex interactions, because these cause larger effects that are harder for selection to alter. This general kind of explanation might apply universally across independent multicellular systems, as supported by the recent finding of the hourglass pattern in plants. We use RNA-seq expression data from the development of the slime mold *Dictyostelium* to demonstrate that it does not follow either of the two canonical patterns but instead tends to show the strongest conservation and weakest evolvability late in development. We propose that this is consistent with a version of the spatial constraints model, modified for organisms that never achieve a high degree of developmental modularity.

Introduction

Multicellularity is one of the major transitions in evolution. It has evolved many times, usually resulting in developmental programs regulating specialization of cells and tissues (Grosberg and Strathmann 2007). The interactions of these specialized cells and tissues both determine the course of development and constrain the pathways it can follow. A major question in development is the extent to which there are common organizational principles that underlie all development (Buss 1987; Gerhart and Kirschner 1997; Schlosser and Wagner 2004a).

One of the oldest generalizations about multicellular development is von Baer's third law which states that early stages of development are most similar among animals, with later stages becoming increasingly divergent (von Baer 1828; Raff 1996). This law figured heavily in the thinking of Darwin and subsequent evolutionary biologists (Gould 1977). More recent work has tended to support a modification, called the hourglass pattern, in which the most constrained stage in animal development is not the earliest stage, but an intermediate one called the phylotypic stage (Sander 1983; Raff 1996). Initially, both von Baer's third law (hereafter called von Baer's law) and the hourglass pattern were simply patterns discovered by morphologists. Two subsequent developments from other fields make them of increasing interest.

First, there are now evo-devo models that attempt to account for these patterns, which suggest that the patterns tell us something fundamental about the way development is structured. The temporal and spatial interactions of development can impose constraints on evolvability owing to the size of their effects (Garfield and Wray 2009) (fig. 1). Both theory (Orr 2000; Otto 2004) and data (Hahn and Kern 2005; He and Zhang

2006; Cooper *et al.* 2007) suggest that genes having large effects or many effects are more conserved because larger changes tend to be more disruptive of adaptation. In development, interactions between stages are expected to constrain early stages, because early changes can have cascading effects later in development (Riedl 1978; Arthur 1988; Schank and Wimsatt 1988). This model, which we call the temporal constraints model, predicts von Baer's third law of conservation of early development. An alternative model focuses on spatial constraints. In this model, spatial interactions within a developmental stage are thought to reach maximum global complexity and constraint at an intermediate developmental stage, with less constraint on earlier stages because the interactions are simpler (Raff *et al.* 1991; Raff 1996; Galis *et al.* 2002). Later stages are assumed to break up into modules, defined as integrated subunits that relatively autonomous and insensitive to outside context (Schlosser and Wagner 2004b), so that interactions become more local with fewer pleiotropic effects. This model is consistent with the hourglass pattern. It is sometimes called the hourglass *model*, but in order to clearly distinguish patterns from models, we will speak of the hourglass *pattern* and the spatial constraints *model*, and this becomes important if, as we will suggest, there is not a one-to-one mapping between models and patterns.

Either of these constraint patterns could be general properties of multicellular development rather than idiosyncratic features of animals, but this question has hardly been explored. A second, more recent development has enabled the study of these patterns at a new level of detail and in new taxa. Molecular methods to study which genes are expressed at different developmental stages, most recently RNA-seq, coupled with evolutionary studies of gene divergence, now allow the study of gene conservation

patterns during development, rather than just morphological patterns. In animals, these new genomic-scale studies of gene expression and evolution confirm the canonical morphological patterns, especially the hourglass model (Hazkani-Covo *et al.* 2005; Artieri *et al.* 2009; Garfield and Wray 2009; Domazet-Loso and Tautz 2010; Kalinka *et al.* 2010; Irie and Kuratani 2011).

These new methods can be applied to any taxon. Because they look across the whole genome, they may be much more powerful than studies of morphology. The first example is the recent finding of support for the hourglass pattern in plants, even though this pattern has apparently never been detected by plant morphologists (Quint *et al.* 2012). This unexpected similarity between animals and plants raises the question of whether the hourglass pattern, and the spatial constraints model underlying it, might be very general or even universal features of multicellular development.

In this study, we explore whether multicellular development in the cellular slime molds or social amoebae follows the hourglass pattern or von Baer's law. The dictyostelids are thought to have evolved multicellularity independently of these two groups (but see Dickinson *et al.* 2012), from single-celled amoebas in the Amoebozoa, the sister taxon to the Opisthokonts (animals and fungi) (King 2004; Grosberg and Strathmann 2007). *Dictyostelium* switches between unicellular and multicellular stages (Kessin 2001). The feeding stage consists of unicellular haploid amoebae. When they run out of food, they send out chemical signals to each other and aggregate into a large mound of cells. This mound ultimately differentiates into a multicellular fruiting body, with some of the cells becoming part of a non-reproductive stalk, and the remainder differentiating as reproductive spores.

Dictyostelium provides a very strong test of the generality of any rules of evolutionary conservation during development because it has a very different developmental program. Instead of developing from a single cell, thousands of unicellular amoebas aggregate before differentiating into a multicellular fruiting body. It can therefore be a chimera of multiple clones (Strassmann *et al.* 2000). It has little cell division during development. Cell fate is determined more by sorting and movement than by initial position (Thompson *et al.* 2004). Rather than many tissues, it has only two main types of cells – spores and supporting stalk – with a handful of stalk cell subtypes (Williams 1997). The stalk is the soma and it comprises only about 20% of the cells.

Despite these differences, early changes should affect later stages, and spatial modules do develop, so either the temporal constraints model or the spatial constraints model might apply. On the other hand, the modules are both few in number and not very independent. Instead of the many compartments of animal development, *Dictyostelium* seems to have only 2 major ones, with several prestalk subtypes (Williams 1997). These compartments remain strongly connected, with stalk and spore cells signaling to each other and moving in relation to each other right up until the fruiting body reaches its final form (Kessin 2001; Dickinson *et al.* 2012). This greatly reduced degree of modularity could change the way that the spatial constraints model applies. Consider the extreme case of development with no real modularity. Then the last half of the hourglass – decreasing constraint at later stages – would never arrive. Instead we would see only the first half in which constraint increases because of increasingly complex global (non-modular) interactions. *Dictyostelium* does not reach this extreme, but is certainly much closer to it than animals and plants.

We compare developmental changes between the model eukaryote *Dictyostelium discoideum* (Eichinger *et al.* 2005) and *D. purpureum* (Sucgang *et al.* 2011), whose proteomes are about as diverged as those of humans and bony fish (Sucgang *et al.* 2011). A recent RNA-seq study provided transcriptional profiles for seven 4–hour time points in both species; it showed significant conservation of gene expression across developmental stages, though with some shift in timing (Parikh *et al.* 2010). We explore four types of conservation for genes expressed at different times in development: the conservation of global expression patterns, the probability of having orthologs, ortholog sequence conservation, and translational selection.

Results

Conservation of Gene Expression Increases with Developmental Time

We first test for conservation of overall expression patterns at different stages, following methods of Irie and Kuratani (Irie and Kuratani 2011). Using the RNA-seq data of Parikh *et al*. (Parikh *et al.* 2010) we correlated gene expression levels for all timepoint comparisons between the two species using the 4743 orthologs with expression data from both. Each correlation is of 4743 gene expression levels at a timepoint in one species with the corresponding expression levels in a timepoint of the other species. Figure 2 shows these correlations, with the highest correlation for each time point (closest match to the other species) marked with open or closed purple circles. If all timepoints are included, the pattern is opposite to the hourglass pattern, with conservation lowest at intermediate stages. Perhaps a fairer test would be to use only the closed purple circles that specify the more truly developmental stages; earlier phases are either vegetative (time 0) or have

expression patterns quite similar to vegetative (Parikh *et al.* 2010) and *D. discoideum* cells become irreversibly committed to development only after 4-6 hours (Katoh *et al.* 2004). With these data, there is a significant increase in conservation of expression pattern with time (fig. 2), the opposite of von Baer's law, with no intermediate maximum.

Sequence Conservation Tends to Increase with Developmental Time

We then tested the effect of developmental time on three measures of sequence conservation. The odds of having orthologs, *O*, measure the probability that genes are present in the other species and that they did not change so rapidly as to fall below the detection threshold. For those genes that do have orthologs, we test the rate of change in aligned non-synonymous sites (dN), subtracted from one (1-dN) to convert it to a measure of conservation (synonymous changes are saturated (Sucgang *et al.* 2011) and therefore provide no useful information). Finally, to add in consideration of insertions and deletions, we calculate the conservation score (CS) for each gene by computing the similarity score for global sequence alignment with its ortholog and re-scaling to a maximum of 1 by dividing by its similarity score against itself (Lopez-Bigas and Ouzounis 2004).

There are relatively few genes with expression specific to one time point (supplementary table S3), so we include all genes and calculate for each a measure of average expression time during development: $t = \sum_i iE_i / \sum_i E_i$, where i is the developmental timepoint and E_i is the expression level of that gene at time i, relative to other genes at that stage. The E_i used is based on the number of RNA-seq reads mapping to a gene during a given stage, but rescaled to adjust for transcript length and for the total number

of reads in a run (Parikh *et al.* 2010). Initially we use time periods i = 8-24 hours in *D. discoideum* and i=12-24 hours in *D. purpureum* because, as noted above, the earlier stages are either vegetative single cells (i = 0) or early aggregation stages that are transcriptionally more similar to the vegetative stage than to later true developmental stages (Parikh *et al.* 2010).

To isolate the effect of timing of expression during development t, we must remove effects from non-developmental causes that could cause bias. For example, high expression level is usually the strongest determinant of sequence conservation (Pal *et al.* 2001; Drummond *et al.* 2005; Wall *et al.* 2005) and genes expressed in more individuals should be also more highly conserved (Van Dyken and Wade 2010). Thus, there is greater conservation of genes expressed in the ubiquitous single-celled vegetative stage. Because these genes are often also expressed early in development (Parikh *et al.* 2010), uncorrected simple linear regressions show that early genes tend to be more conserved, as in von Baer's law (supplementary table S1). Similarly, there should be greater conservation of genes expressed across all stages, which necessarily have an intermediate average expression time t, with the result that quadratic regressions show humped curves (negative t^2 coefficients), as in the hourglass pattern (supplementary table S1). To remove such non-developmental effects, we run multiple regressions including vegetative expression v (= E_0 above) and an index of stage specificity of expression τ. We also include the average expression level across the developmental timepoints d and gene length l as covariates.

For all three measures of conservation the t^2 terms in multiple regressions were insignificant (Table 1), providing no support for an hourglass pattern. We therefore shift

focus to linear regressions without the quadratic term. For the odds of having an ortholog, logistic multiple regression showed *t* coefficients to be positive ($p < 0.001$), indicating late conservation, the opposite of von Baer's law. The same conclusion holds for 1-dN and CS using *D. purpureum* expression data (both $p < 0.001$) though the coefficients are not significant when using *D. discoideum* (Table 1, supplementary table S2).

The lack of fit to both von Baer's law and the hourglass pattern remains when we redo the above analyses using all developmental stages (supplementary table S3), or using only the genes that are most specific to development (supplementary table S4, fig. S1), or to individual timepoints (supplementary table S5), or to somatic (pre-stalk) tissues (supplementary table S6 and fig. S2). Among these 29 new quadratic regressions there was only one significant negative t^2 effect supporting an hourglass pattern, and no negative effects of *t* in the linear regressions. Instead most of the *t* effects are significantly positive, confirming and strengthening our initial finding of late constraint.

Translational Selection Does Not Support Von Baer's Law

One further pattern reinforces the general message that temporal patterns of constraint are different in *Dictyostelium* compared to animals. Gene sequence conservation is generally highly correlated with expression level (Pal *et al.* 2001; Drummond *et al.* 2005; Wall *et al.* 2005), reflecting selection for accurate translation, perhaps because of a greater importance of proper protein folding in highly expressed genes (Drummond *et al.* 2005; Drummond and Wilke 2008). If translational or misfolding errors early in development are more serious because of cascading effects later, then we would expect sequence

conservation to be more related to gene expression level early in development than later in development. That pattern is not found. Instead the strength of the relationship for *D. discoideum* (again corrected for vegetative expression level *v*) peaks at intermediate values (fig. 3a). This supports the hourglass pattern, but it is not a robust finding. It is contradicted by strictly increasing trends when the same analysis is done using *D. purpureum* as the focal species (fig. 3b, supplementary table S7), and when using only the most developmental genes in either species (supplementary fig. S3). Again the predominant pattern indicates late conservation.

The timing of expression and developmental defects

Given that neither the von Baer nor the hourglass patterns seems to hold, it is worth asking whether the assumptions of their underlying models are met. The temporal constraint model assumes that early changes have larger and therefore more deleterious effects, while the spatial constraint model assumes that more deleterious changes occur at some at intermediate stage. A list of genes whose mutants have known developmental defects is maintained by Dictybase (http://dictybase.org/Downloads/). For genes with stage-specific expression (2-fold higher expression at that stage than at any other), we calculated the proportion that appear on the list of genes with developmental defects (fig. 4). This fraction was highest early in development and declined at later stages, which accords best with the assumption of the temporal model.

Discussion

The problem of the similarity and difference of developmental stages goes back to the early 19th century with von Baer's studies (von Baer 1828). Von Baer's law became an important part of early evolutionary theory but thinking about the relationship between development and evolution receded after the rise of genetics (Gould 1977). Yet the problem of how multicellular developmental programs are constrained remains an important one, now linked to models of network interactions and pleiotropy. A temporal constraints model positing that early changes have pleiotropic later effects, but not vice versa, is consistent with von Baer's law (Riedl 1978; Arthur 1988; Schank and Wimsatt 1988). A model assuming that spatial constraints are more important is consistent with the hourglass pattern (Raff *et al.* 1991; Raff 1996), assuming that modularity late in development reduces constraints by making most interactions local rather than global. On the empirical side, the development of RNA-seq methods now allow us to thoroughly catalog the genes being expressed during different developmental stages, so we can conduct powerful studies of conservation of all genes rather than of a modest number of morphological traits. The recent application of these techniques to plants showed support for the hourglass pattern, even though that pattern has not been evident from morphological studies (Quint *et al.* 2012). This raises the possibility that the hourglass pattern might be common to many or all multicellular systems, perhaps reflecting widespread conformity to the spatial constraints model.

We find however that *Dictyostelium* development does not follow either of the canonical animal patterns. Instead, most analyses point to a novel pattern: greater conservation of the genes expressed later in development. We can therefore reject both the temporal constraint model and the standard spatial constraint model, at least as they

have been conceived up to this point. However the possible relationships between patterns and process may be more complex than has been thought. The rejection of the temporal model seems strong but the spatial constraint model, suitably modified, could account for results. Our predominant result resembles the first half of the hourglass; conservation increases with time but then does not show the decrease of the second half of the hourglass. Perhaps *Dictyostelium* simply never reaches the second half, where constraints are reduced by a break-up into relatively independent modules. Though prespore and prestalk regions do separate relatively early, the two tissue types continue to interact with each other, and to move with respect to each other, throughout the rest of development. The final spatial relationship is not achieved until the culmination of development, when the prespore cells, following signals from stalk cells in the tip, move up to the top of the stalk (Dickinson *et al.* 2012). Final maturation of spore cells is triggered in part by signals from stalk cells (Wang *et al.* 1999).

Therefore, we propose that the spatial model be divided into two: the existing spatial model with strong modularity that predicts the hourglass pattern, and a new spatial model for organisms with little or no modularity, which predicts a new pattern of increasing constraint throughout development. By analogy with the hourglass pattern, von Baer's law is sometimes called the funnel pattern (Irie and Kuratani 2011), and our new pattern could be called the inverted funnel. Table 2 shows a classification of models, patterns, and their relationship. Though our results could be viewed as a rejection of what has been called the hourglass model (the spatial constraints model), they can also be viewed as providing a generalization and strengthening of the basic logic that underlies that model, provided the assumption of late modularity is revised. Theories are best tested not

by repeated confirmation of the same prediction but by novel predictions arising in special cases. The most basic feature of the thinking behind the hourglass model is neither the hourglass pattern itself nor the assumptions of a particular form of development. Instead the main feature is the idea that the time of maximum conservation will correspond to the time of maximum global interaction. A difference in the time of maximum global interaction should cause a difference in the time of maximum conservation, and that is what we appear to see in *Dictyostelium*.

Our analysis showing that genes expressed earlier in development are more likely to have developmental defects (fig 4) weighs against this interpretation. This pattern is more consistent with the assumptions of the temporal model. The *prima facie* predictions for the spatial models with and without modules would be more defects for genes with intermediate and late expression times, respectively. There may be biases in the dataset of mutant defects, because late development has been less heavily studied (Dickinson *et al.* 2012), or because severe late defects such as defective spores may be less noticeable, but the analysis at least suggests that we should be receptive to alternative explanations.

As noted in the introduction, *Dictyostelium* development differs in many ways from development in animals. The smaller soma (stalk) does not seem to be the explanation for the failure to find the canonical patterns, because they also fail to appear when we examine prestalk genes only (supplementary fig. S2, table S6). However, it is impossible at this stage to rule out roles of some of the other differences in *Dictyostelium*'s development. For example, since development occurs after aggregation, the multicellular stage can contain multiple clones, which compete to become spores instead of dead stalk. Evolutionary conflict can lead to continual adaptive evolution and increased evolutionary

rates (Nielsen 2005) and this might explain higher evolutionary rates early in *Dictyostelium* development if competition is strongest at the earlier stages when prespore and prestalk roles are initially assigned. However, competition between clones within fruiting bodies may be relatively infrequent, as the only study of genetic structure in the field shows that most fruiting bodies consist of a single clone (Gilbert *et al.* 2007).

Given all the developmental differences between *Dictyostelium* and animals, it may prove difficult to isolate low modularity as the key to its pattern of late conservation. Gene ontology patterns can identify developmental categories of genes that are expressed at different times (Parikh *et al.* 2010) but they do not readily distinguish global from more local interactions. *Dictyostelium* developmental biologists should use their expertise on the details of development to weigh in on whether the period of maximum global interaction is early, intermediate, or late. But perhaps the best way forward would be to test the model in other organisms with low modularity. Simple animals like corals or *Trichoplax* might fit the model. So might simpler plants, such as fern gametophytes or even moss sporophytes. Fungal fruiting bodies are another possibility. If the model proves successful, it might be used as an indicator or the degree of complex modularity in development. For example, we would expect to see the non-modular late-conservation pattern in the simpler multicellular members of the Volvocales but, if they are sufficiently complex, the most multicellular forms like *Volvox* would show the hourglass pattern.

Materials and Methods

Genome and Transcriptome Data. We retrieved predicted protein-coding gene models for *D. discoideum* (version 03-10-2010) and *D. purpureum* (version 02-04-2010) from

dictyBase (http://dictybase.org/db/cgi-bin/dictyBase/download/blast_databases.pl). Orthologous pairs were found using reciprocal best matches of the BLASTP using the Inparanoid algorithm (Ostlund *et al.* 2010), as previously described (Sucgang *et al.* 2011). Parikh and colleagues provided normalized RNA-seq transcriptome data at 4-hour intervals and from prestalk and prespore cells (http://dictygenome.bcm.tmc.edu/~anup/rnaseq). For each species, we averaged the RNA-seq read counts for each gene from the two biological replicates of each timepoint. We eliminated genes that do not have a stage with at least 30 reads from at least one timepoint (equal to the length of average one transcript) and genes that are not biologically reproducible in the RNA-seq replicates. This resulted in 5259 *D. discoideum* genes and 6014 *D. purpureum* genes with othologs in the other species.

Computational and Statistical Analysis. We estimated nonsynonymous nucleotide substitution rates (dN) by the maximum likelihood program *codeml* of PAML4 (Yang 2007), based on retro-translated protein sequence alignments from GAP4 (Huang and Brutlag 2007) global alignments (synonymous changes are saturated (Sucgang *et al.* 2011)). We calculated conservation scores (CS) (Lopez-Bigas and Ouzounis 2004), which range from 0 to 1, from GAP4 global alignments.

To estimate stage specificity of expression, we applied the tissue specificity index τ (Yanai *et al.* 2005) to our 7 stages instead of to tissues. We normalized variables as following: *dN*, square root log transform; *CS* and τ, arcsine square root transform; gene length and RNA-seq read count, log transform. We used the R modules *glm* to construct

logistic models, and the module *lm* for linear models (R Development Core Team 2005). We also used R modules *car* and *effects*.

Supplementary Material

Supplementary figures and tables are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Gad Shaulsky for access to RNA-seq data and referees for helpful comments. This material is based upon work supported by the National Science Foundation under grants DEB-0918931 and DEB-1146375.

Tables

Table 1. Sequence conservation for genes expressed at different developmental times.

Conservation measure [a]	Quadratic regression t^2 coeff. [c]		Linear regression t coeff.	
	D. discoideum	*D. purpureum*	*D. discoideum*	*D. purpureum*
O	0.0050 ns [b]	-0.0060 ns	0.039 ***	0.63 ***
CS	0.00029 ns	0.00097 ns	0.00071 ns	0.0061 ***
1-*dN*	0.00014 ns	0.00054 ns	0.00073 ns	0.0035 ***

[a] *O* = odds of having an ortholog, logistic regression; *CS* = protein sequence conservation score, linear regression; 1-*dN* = 1- nonsynonymous substitution rate; linear regression.

[b] *** coefficient significant, ns = not significant.

[c] A negative t^2 coefficient tests is required to support the hourglass model; a negative t coefficient (in a linear regression lacking a t^2 term) supports von Baer's law. Full regressions and gene numbers are given in Table S2.

Table 2. Models and patterns of evolutionary conservation during development

Model	Predicted pattern	Stage of maximum conservation
temporal	Von Baer's law (funnel)	early
spatial constraints with modules	hourglass pattern	intermediate
spatial constraints without modules	first half of hourglass; (inverse funnel)	late

Figure Legends

Fig. 1. Schematic of development, with proposed effects on conservation. Dashed grey arrows show that early stages affect later stages, so that early changes create large effects that are less likely to be favored by selection, leading to von Baer's law (dashed grey plot). Solid black arrows indicate spatial interactions, which may be most complex at the middle stage, when complex global interactions set up spatial modules or compartments, represented as four sub-regions. Early stages have simpler global interactions and late ones are mainly modular, yielding the hourglass pattern of conservation (solid black plot).

Fig. 2. Expression is more conserved later in development. Expression similarity was calculated as Pearson's correlation coefficients of normalized gene expression levels between 4-hr timepoints between the two species (Parikh *et al.* 2010). Each correlation is an average over the four pairwise comparisons of the two biologically-replicated transcriptomes from each species timepoint. The plot follows those of Irie and Kuratani (Irie and Kuratani 2011) with (**a**) *D. discoideum* timepoints on the x-axis and *D. purpureum* timepoints shown as colored lines and (**b**) vice versa. The purple circles show the highest correlation for each timepoint, pointing to the most similar timepoint of the other species. Linear regression on the solid purple points shows this measure of expression conservation increases significantly with development time i (**a**, $R^2 = 0.82$, $p = 0.03$; **b**, $R^2 = 0.97$, $p = 0.02$). Pre-developmental stages are shown as open circles and dashed lines.

Fig. 3. Relative translational selection at different stages. For each time point of (**a**) *D. discoideum* and (**b**) *D. purpureum*, we regressed gene sequence conservation (*CS* or 1-*dN*) on gene expression level (rescaled RNA-seq read counts) at that time, partialing out the effects of level of vegetative expression level (regression data for each timepoint shown in table S7). Higher translational (purifying) selection at a timepoint should be reflected in a higher partial regression coefficient for that timepoint. The error bars show standard errors. To show the pattern with expression time (dashed lines), we perform another regression, of the partial coefficients on expression time i (**a**. CS, $p = 0.01$ for both i^2 and i coefficients; 1-dN, $p = 0.001$ for both. **b**. i^2 terms not sig., so linear regressions are shown; CS, p=0.04; 1-dN, p=0.1).

Fig. 4. The proportion of stage-specific genes associated with known developmental defects. The stage-specific genes are those with at least 2-fold higher number of RNA-seq reads at that stage than at nay other stage, including the vegetative stage (N = 65, 26, 105, 99, 320). For stages 8-24, the proportion of these genes matching a list of genes with known developmental defects (N = 1,015, http://dictybase.org/Downloads/, 16-03-2011) is plotted.

References

Arthur W. 1988. A theory of the evolution of development. Chichester: John Wiley and Sons.

Artieri CG, Haerty W, Singh RS 2009. Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of *Drosophila*. *BMC Biology* 7: 42.

Buss LW 1987. The evolution of individuality. Princeton: Princeton University Press.

Cooper TF, Ostrowski EA, Travisano M 2007. A negative relationship between mutation pleiotropy and fitness effect in yeast. *Evolution* 61: 1495-1499. doi: 10.1111/j.1558-5646.2007.00109.x

Dickinson DJ, Nelson WJ, Weis WI 2012. An epithelial tissue in *Dictyostelium* challenges the traditional origin of metazoan multicellularity. *Bioessays* 34: 833-840.

Domazet-Loso T, Tautz D 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468: 815-819. doi: 10.1038/nature09632

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH 2005. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* 102: 14338-14343. doi: 10.1073/pnas.0504070102

Drummond DA, Wilke CO 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134: 341-352. doi: 10.1016/j.cell.2008.05.042

Eichinger L, Pachebat JA, Glöckner G, et al. (98 co-authors) 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435: 43-57.

Galis F, van Dooren TJM, Metz JAJ 2002. Conservation of the segmented germband stage: robustness or pleiotropy. *Trends Genet.* 18: 504-509.

Garfield DA, Wray GA 2009. Comparative embryology without a microscope: using genomic approaches to understand the evolution of development. *J Biol.* 8: 65.

Gerhart G, Kirschner M. 1997. Cells, embryos and evolution; toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability. Malden (MA): Blackwell Science.

Gilbert OM, Foster KR, Mehdiabadi NJ, Strassmann JE, Queller DC 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc Natl Acad Sci U S A* 104: 8913-8917.

Gould SJ. 1977. Ontogeny and phylogeny. Cambridge (MA): Harvard University Press.

Grosberg RK, Strathmann RR 2007. The evolution of multicellularity: A minor major transition? *Annu Rev Ecol Evol Syst.* 38: 621-654. doi: 10.1146/annurev.ecolsys.36.102403.114735

Hahn MW, Kern AD 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol.* 22: 803-806. doi: 10.1093/molbev/msi072

Hazkani-Covo E, Wool D, 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 Exptl Zool*. 304B: 150-158.

He XL, Zhang JZ 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173: 1885-1891. doi: 10.1534/genetics.106.060269

Huang S, Brutlag DL 2007. Dynamic use of multiple parameter sets in sequence alignment. *Nucl Acids Res*. 35: 678-686.

Irie N, Kuratani S 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Commun*. 2: 248.

Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature* 468: 811-814. doi: 10.1038/nature09634

Katoh M, Shaw C, Xu QK, VanDriessche N, Morio T, Kuwayama H, Obara S, Urushihara H, Tanaka Y, Shaulsky G 2004. An orderly retreat: dedifferentiation is a regulated process. *Proc. Natl. Acad. Sci. U S A* 101: 7005-7010.

Kessin RH. 2001. Dictyostelium: evolution, cell biology, and the development of multicellularity. Cambridge UK: Cambridge University Press.

King N 2004. The unicellular ancestry of animal development. *Devel Cell* 7: 313-325.

Lopez-Bigas N, Ouzounis CA 2004. Genome-wide identification of genes likely to be involved in human genetic disease. *Nucl Acids Res*. 32: 3108-3114.

Nielsen R 2005. Molecular signatures of natural selection. A*nnu Rev Genet*. 39: 197-218.

Orr HA 2000. Adaptation and the cost of complexity. *Evolution* 54: 13-20. doi: 10.1111/j.0014-3820.2000.tb00002.x

Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O, Sonnhammer ELL 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucl Acids Res*. 38: D196-D203. doi: 10.1093/nar/gkp931

Otto S 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc Roy Soc B* 271: 705-714.

Pal C, Papp B, Hurst LD 2001. Highly expressed genes in yeast evolve slowly. *Genetics* 158: 927-931.

Parikh A, Miranda ER, Katoh-Kurasawa M, et al. (13 co-authors) 2010. Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biology* 11: R35. doi: 10.1186/gb-2010-11-3-r35

Quint M, Drost H-G, Gabel A, Ullrich KK, Bönn M, Grosse I 2012. A trancriptomic hourglass in plant embryogenesis. *Nature* 490: 98-101.

R Development Core Team. 2005. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raff R, Wray G, Henry J. 1991. Implications of radical evolutionary change in early development for concepts of developmental constraint. In: Warren L, Koprowski H, editors. New perspectives on evolution. New York: Wiley-Liss. p. 189-207.

Raff RA. 1996. The shape of life: genes, development, and the evolution of animal form. Chicago: University of Chicago Press.

Riedl R. 1978. Order in living organisms: a systems analysis of evolution. New York: John Wiley & Sons.

Sander K. 1983. The evolution of patterning mechanisms: gleanings from insect embryogenesis and spermatogenesis. In: Goodwin B, Holder N, Wylie C, editors. Development and evolution. Cambridge (MA): Cambridge University Press. p. 137-159.

Schank JC, Wimsatt WC. 1988. Generative entrenchment and evolution. In: Fine A, Machamen PK, editors. PSA-1986 Vol II: Philosophy of Science Association. p. 33-60.

Schlosser G, Wagner GP (editors). 2004a. Modularity in development and evolution. In. Chicago: University of Chicago Press. p. 600.

Schlosser G, Wagner G. 2004b. Introduction: the modularity concept in developmental and evolutionary biology, In: Schlosser G, Wagner GP, editors. Modularity in development and evolution. Chicago: University of Chicago Press, pp 1-11.

Strassmann JE, Zhu Y, Queller DC 2000. Altruism and social cheating in the social amoeba, *Dictyostelium discoideum*. *Nature* 408: 965-967.

Sucgang R, Kuo A, Tian XJ, (43 co-authors) 2011. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum. Genome Biology* 12: R20. doi: 10.1186/gb-2011-12-2-r20

Thompson CRL, Reichelt S, Kay RR 2004. A demonstration of pattern formation without positional information in *Dictyostelium*. *Devel Growth Diff* 46: 363-369.

Van Dyken JD, Wade MJ 2010. The genetic signature of conditional expression. *Genetics* 184: 557-570.

von Baer KE. 1828. Über Entwicklungsgeschichte der Thiere: Beobachtung und Reflexion. Königsberg: Gebrüder Bornträger.

Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, Feldman MW 2005. Functional genomic analysis of the rates of protein evolution. *Proc Natl Acad Sci U S A* 102: 5483-5488. doi: 10.1073/pnas.0501761102

Wang N, Soderbom F, Anjard C, Shaulsky G, Loomis WF 1999. SDF-2 induction of terminal differentiation in *Dictyostelium discoideum* is mediated by the membrane-spanning sensor kinase DhkA. Mol. Cell. Biol. 19: 4750-4756.

Williams J. 1997. Prestalk and stalk heterogeneity in *Dictyostelium*. In: Maeda Y, Inouye K, Takeuchi I, editors. *Dictyostelium* - A model system for cell and developmental biology. Tokyo, Japan: Universal Academy Press. p. 293-304.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, Lancet D, Shmueli O 2005. Genome-wide midrange *transcription* profiles reveal expression level relationships in human tissue specification. Bioinformatics 21: 650-659. doi: 10.1093/bioinformatics/bti042

Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and *Evolution* 24: 1586-1591.

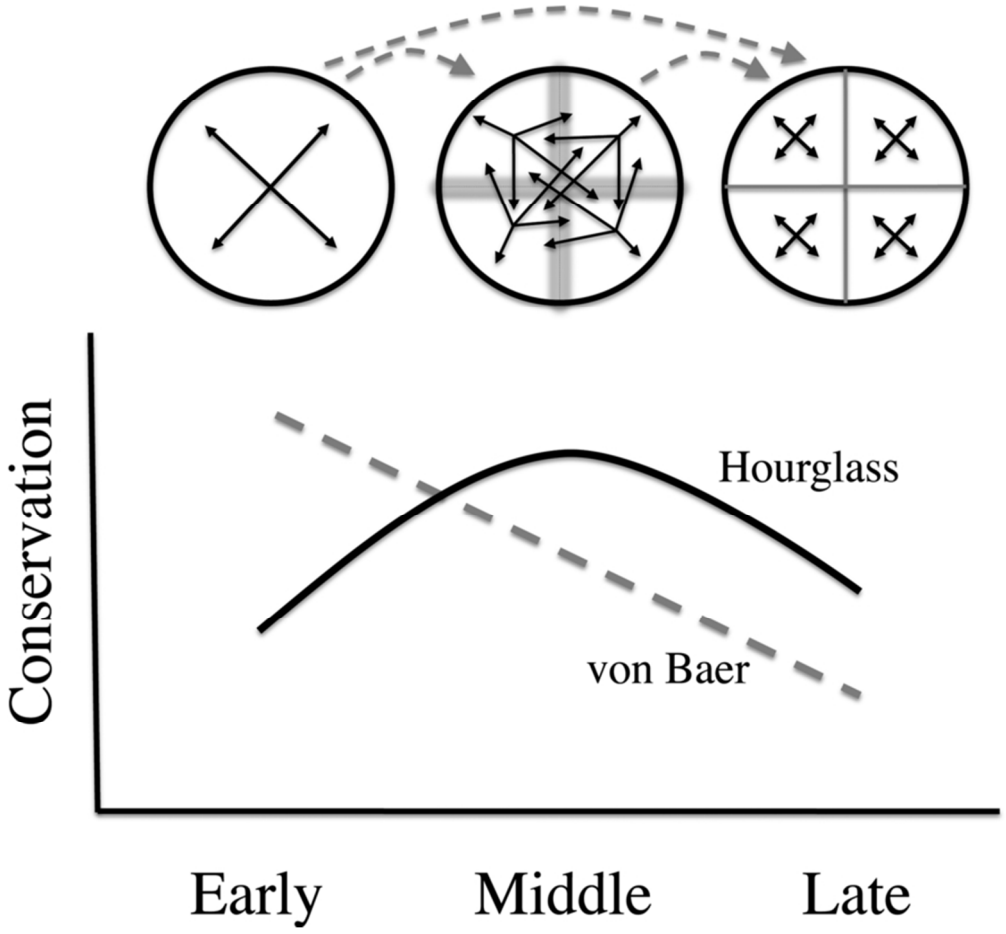

Fig. 1. Schematic of development, with proposed effects on conservation. Dashed grey arrows show that early stages affect later stages, so that early changes create large effects that are less likely to be favored by selection, leading to von Baer's law (dashed grey plot). Solid black arrows indicate spatial interactions, which may be most complex at the middle stage, when complex global interactions set up spatial modules or compartments, represented as four sub-regions. Early stages have simpler global interactions and late ones are mainly modular, yielding the hourglass pattern of conservation (solid black plot).
75x70mm (300 x 300 DPI)

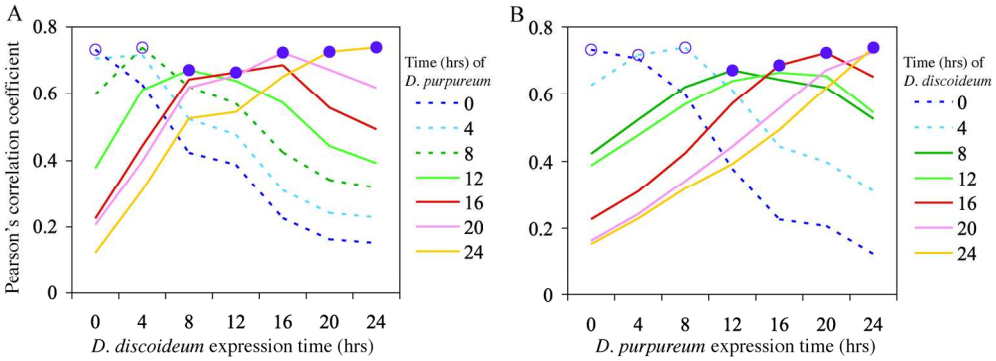

Fig. 2. Expression is more conserved later in development. Expression similarity was calculated as Pearson's correlation coefficients of normalized gene expression levels between 4-hr timepoints between the two species (Parikh et al. 2010). Each correlation is an average over the four pairwise comparisons of the two biologically-replicated transcriptomes from each species timepoint. The plot follows those of Irie and Kuratani (Irie and Kuratani 2011) with (a) D. discoideum timepoints on the x-axis and D. purpureum timepoints shown as colored lines and (b) vice versa. The purple circles show the highest correlation for each timepoint, pointing to the most similar timepoint of the other species. Linear regression on the solid purple points shows this measure of expression conservation increases significantly with development time i (a, R^2 = 0.82, p = 0.03; b, R^2 = 0.97, p = 0.02). Pre-developmental stages are shown as open circles and dashed lines.
171x62mm (300 x 300 DPI)

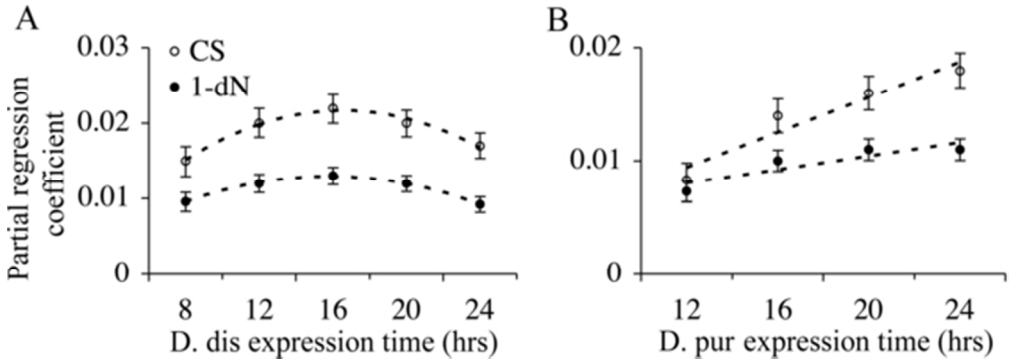

Fig. 3. Relative translational selection at different stages. For each time point of (a) D. discoideum and (b) D. purpureum, we regressed gene sequence conservation (CS or 1-dN) on gene expression level (rescaled RNA-seq read counts) at that time, partialing out the effects of level of vegetative expression level (regression data for each timepoint shown in table S7). Higher translational (purifying) selection at a timepoint should be reflected in a higher partial regression coefficient for that timepoint. The error bars show standard errors. To show the pattern with expression time (dashed lines), we perform another regression, of the partial coefficients on expression time i (a. CS, p = 0.01 for both i^2 and i coefficients; 1-dN, p = 0.001 for both. b. i^2 terms not sig., so linear regressions are shown; CS, p=0.04; 1-dN, p=0.1). 54x19mm (300 x 300 DPI)

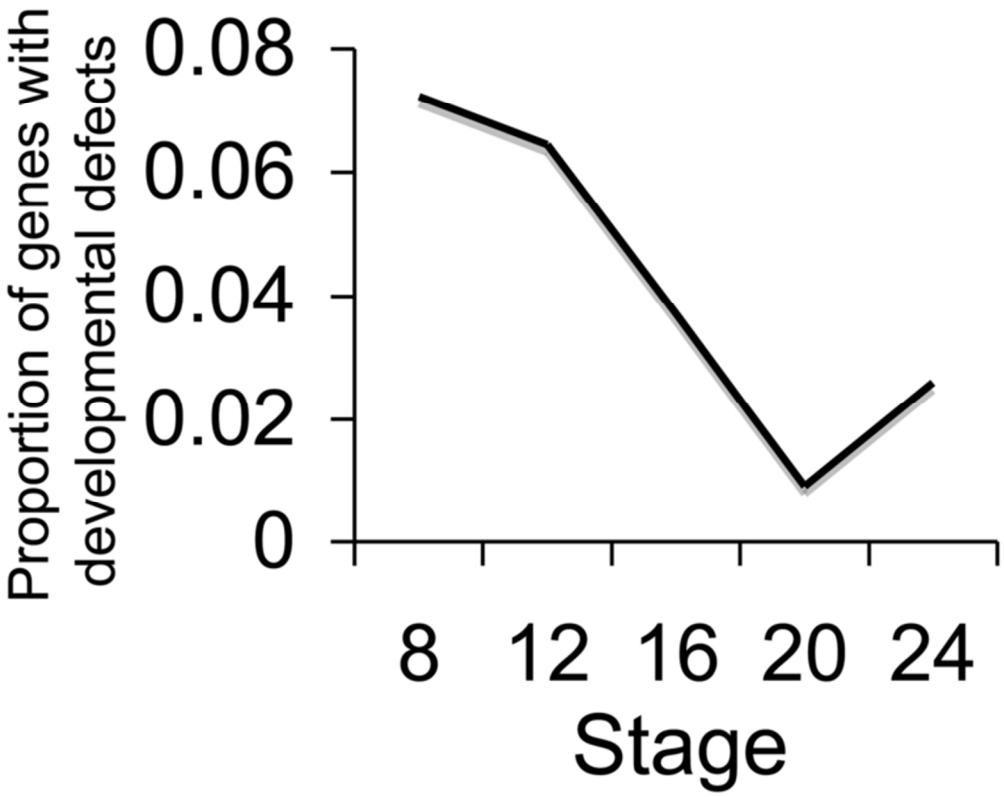

Fig. 4. The proportion of stage-specific genes associated with known developmental defects. The stage-specific genes are those with at least 2-fold higher number of RNA-seq reads at that stage than at nay other stage, including the vegetative stage (N = 65, 26, 105, 99, 320). For stages 8-24, the proportion of these genes matching a list of genes with known developmental defects (N = 1,015, http://dictybase.org/Downloads/, 16-03-2011) is plotted.
64x51mm (300 x 300 DPI)