

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

Follow this and additional works at: http://openscholarship.wustl.edu/bio_facpubs

 Part of the [Biology Commons](#)

Recommended Citation

Tian, Xiangjun; Strassmann, Joan E.; and Queller, David C., "Dictyostelium development shows a novel pattern of evolutionary conservation" (2013). *Biology Faculty Publications & Presentations*. Paper 31.

http://openscholarship.wustl.edu/bio_facpubs/31

This Article is brought to you for free and open access by the Biology at Washington University Open Scholarship. It has been accepted for inclusion in Biology Faculty Publications & Presentations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.



Dictyostelium development shows a novel pattern of evolutionary conservation

Journal:	<i>Molecular Biology and Evolution</i>
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

SCHOLARONE™
Manuscripts

1
2
3 ***Dictyostelium* development shows a novel pattern of evolutionary**
4
5
6 **conservation**
7

8
9 The submission is intended as an Article

10
11 Xiangjun Tian, Joan E. Strassmann, and David C. Queller¹
12

13
14 Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA
15

16 ¹Corresponding author:
17

18 David C. Queller
19 Department of Biology CB 1137
20 Washington University in St. Louis
21 One Brookings Drive
22 St. Louis MO 63130 U.S.A.
23 Tel: (+1) 314-935-3528
24 Fax: (+1) 314-935-4432
25 Email: queller@wustl.edu
26
27

28
29 **Key words:** von Baer's law, hourglass model, *Dictyostelium*, evolutionary conservation,
30
31 development
32

33
34 **Running head:** Evolutionary conservation of *Dictyostelium* development
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 2006; Cooper *et al.* 2007) suggest that genes having large effects or many effects are
4 more conserved because larger changes tend to be more disruptive of adaptation. In
5
6 development, interactions between stages are expected to constrain early stages, because
7
8 early changes can have cascading effects later in development (Riedl 1978; Arthur 1988;
9
10 Schank and Wimsatt 1988). This model, which we call the temporal constraints model,
11
12 predicts von Baer's third law of conservation of early development. An alternative model
13
14 focuses on spatial constraints. In this model, spatial interactions within a developmental
15
16 stage are thought to reach maximum global complexity and constraint at an intermediate
17
18 developmental stage, with less constraint on earlier stages because the interactions are
19
20 simpler (Raff *et al.* 1991; Raff 1996; Galis *et al.* 2002). Later stages are assumed to break
21
22 up into modules, defined as integrated subunits that relatively autonomous and insensitive
23
24 to outside context (Schlosser and Wagner 2004b), so that interactions become more local
25
26 with fewer pleiotropic effects. This model is consistent with the hourglass pattern. It is
27
28 sometimes called the hourglass *model*, but in order to clearly distinguish patterns from
29
30 models, we will speak of the hourglass *pattern* and the spatial constraints *model*, and this
31
32 becomes important if, as we will suggest, there is not a one-to-one mapping between
33
34 models and patterns.
35
36
37
38
39
40
41
42

43
44 Either of these constraint patterns could be general properties of multicellular
45
46 development rather than idiosyncratic features of animals, but this question has hardly
47
48 been explored. A second, more recent development has enabled the study of these
49
50 patterns at a new level of detail and in new taxa. Molecular methods to study which
51
52 genes are expressed at different developmental stages, most recently RNA-seq, coupled
53
54 with evolutionary studies of gene divergence, now allow the study of gene conservation
55
56
57
58
59
60

1
2
3 patterns during development, rather than just morphological patterns. In animals, these
4
5 new genomic-scale studies of gene expression and evolution confirm the canonical
6
7 morphological patterns, especially the hourglass model (Hazkani-Covo *et al.* 2005;
8
9 Artieri *et al.* 2009; Garfield and Wray 2009; Domazet-Lošo and Tautz 2010; Kalinka *et*
10
11 *al.* 2010; Irie and Kuratani 2011).

12
13
14
15 These new methods can be applied to any taxon. Because they look across the
16
17 whole genome, they may be much more powerful than studies of morphology. The first
18
19 example is the recent finding of support for the hourglass pattern in plants, even though
20
21 this pattern has apparently never been detected by plant morphologists (Quint *et al.*
22
23 2012). This unexpected similarity between animals and plants raises the question of
24
25 whether the hourglass pattern, and the spatial constraints model underlying it, might be
26
27 very general or even universal features of multicellular development.

28
29
30
31
32 In this study, we explore whether multicellular development in the cellular slime
33
34 molds or social amoebae follows the hourglass pattern or von Baer's law. The
35
36 dictyostelids are thought to have evolved multicellularity independently of these two
37
38 groups (but see Dickinson *et al.* 2012), from single-celled amoebas in the Amoebozoa,
39
40 the sister taxon to the Opisthokonts (animals and fungi) (King 2004; Grosberg and
41
42 Strathmann 2007). *Dictyostelium* switches between unicellular and multicellular stages
43
44 (Kessin 2001). The feeding stage consists of unicellular haploid amoebae. When they
45
46 run out of food, they send out chemical signals to each other and aggregate into a large
47
48 mound of cells. This mound ultimately differentiates into a multicellular fruiting body,
49
50 with some of the cells becoming part of a non-reproductive stalk, and the remainder
51
52 differentiating as reproductive spores.
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1
2
3 We compare developmental changes between the model eukaryote *Dictyostelium*
4 *discoideum* (Eichinger *et al.* 2005) and *D. purpureum* (Sucgang *et al.* 2011), whose
5
6 proteomes are about as diverged as those of humans and bony fish (Sucgang *et al.* 2011).
7
8 A recent RNA-seq study provided transcriptional profiles for seven 4-hour time points in
9
10 both species; it showed significant conservation of gene expression across developmental
11
12 stages, though with some shift in timing (Parikh *et al.* 2010). We explore four types of
13
14 conservation for genes expressed at different times in development: the conservation of
15
16 global expression patterns, the probability of having orthologs, ortholog sequence
17
18 conservation, and translational selection.
19
20
21
22
23
24
25
26

27 Results

28 Conservation of Gene Expression Increases with Developmental Time

29
30 We first test for conservation of overall expression patterns at different stages, following
31
32 methods of Irie and Kuratani (Irie and Kuratani 2011). Using the RNA-seq data of Parikh
33
34 *et al.* (Parikh *et al.* 2010) we correlated gene expression levels for all timepoint
35
36 comparisons between the two species using the 4743 orthologs with expression data from
37
38 both. Each correlation is of 4743 gene expression levels at a timepoint in one species with
39
40 the corresponding expression levels in a timepoint of the other species. Figure 2 shows
41
42 these correlations, with the highest correlation for each time point (closest match to the
43
44 other species) marked with open or closed purple circles. If all timepoints are included,
45
46 the pattern is opposite to the hourglass pattern, with conservation lowest at intermediate
47
48 stages. Perhaps a fairer test would be to use only the closed purple circles that specify the
49
50 more truly developmental stages; earlier phases are either vegetative (time 0) or have
51
52
53
54
55
56
57
58
59
60

1
2
3 expression patterns quite similar to vegetative (Parikh *et al.* 2010) and *D. discoideum*
4
5 cells become irreversibly committed to development only after 4-6 hours (Katoh *et al.*
6
7 2004). With these data, there is a significant increase in conservation of expression
8
9 pattern with time (fig. 2), the opposite of von Baer's law, with no intermediate maximum.
10
11

12 13 14 15 **Sequence Conservation Tends to Increase with Developmental Time**

16
17 We then tested the effect of developmental time on three measures of sequence
18
19 conservation. The odds of having orthologs, O , measure the probability that genes are
20
21 present in the other species and that they did not change so rapidly as to fall below the
22
23 detection threshold. For those genes that do have orthologs, we test the rate of change in
24
25 aligned non-synonymous sites (dN), subtracted from one (1-dN) to convert it to a
26
27 measure of conservation (synonymous changes are saturated (Sucgang *et al.* 2011) and
28
29 therefore provide no useful information). Finally, to add in consideration of insertions
30
31 and deletions, we calculate the conservation score (CS) for each gene by computing the
32
33 similarity score for global sequence alignment with its ortholog and re-scaling to a
34
35 maximum of 1 by dividing by its similarity score against itself (Lopez-Bigas and
36
37 Ouzounis 2004).
38
39
40
41
42

43
44 There are relatively few genes with expression specific to one time point
45
46 (supplementary table S3), so we include all genes and calculate for each a measure of
47
48 average expression time during development: $t = \sum_i i E_i / \sum_i E_i$, where i is the developmental
49
50 timepoint and E_i is the expression level of that gene at time i , relative to other genes at
51
52 that stage. The E_i used is based on the number of RNA-seq reads mapping to a gene
53
54 during a given stage, but rescaled to adjust for transcript length and for the total number
55
56
57
58
59
60

1
2
3 of reads in a run (Parikh *et al.* 2010). Initially we use time periods $i = 8-24$ hours in *D.*
4
5 *discoideum* and $i=12-24$ hours in *D. purpureum* because, as noted above, the earlier
6
7 stages are either vegetative single cells ($i = 0$) or early aggregation stages that are
8
9 transcriptionally more similar to the vegetative stage than to later true developmental
10
11 stages (Parikh *et al.* 2010).
12
13

14
15 To isolate the effect of timing of expression during development t , we must remove
16
17 effects from non-developmental causes that could cause bias. For example, high
18
19 expression level is usually the strongest determinant of sequence conservation (Pal *et al.*
20
21 2001; Drummond *et al.* 2005; Wall *et al.* 2005) and genes expressed in more individuals
22
23 should be also more highly conserved (Van Dyken and Wade 2010). Thus, there is
24
25 greater conservation of genes expressed in the ubiquitous single-celled vegetative stage.
26
27 Because these genes are often also expressed early in development (Parikh *et al.* 2010),
28
29 uncorrected simple linear regressions show that early genes tend to be more conserved, as
30
31 in von Baer's law (supplementary table S1). Similarly, there should be greater
32
33 conservation of genes expressed across all stages, which necessarily have an intermediate
34
35 average expression time t , with the result that quadratic regressions show humped curves
36
37 (negative t^2 coefficients), as in the hourglass pattern (supplementary table S1). To remove
38
39 such non-developmental effects, we run multiple regressions including vegetative
40
41 expression v ($= E_0$ above) and an index of stage specificity of expression τ . We also
42
43 include the average expression level across the developmental timepoints d and gene
44
45 length l as covariates.
46
47
48
49
50
51
52

53 For all three measures of conservation the t^2 terms in multiple regressions were
54
55 insignificant (Table 1), providing no support for an hourglass pattern. We therefore shift
56
57
58
59
60

1
2
3 focus to linear regressions without the quadratic term. For the odds of having an
4 ortholog, logistic multiple regression showed t coefficients to be positive ($p < 0.001$),
5 indicating late conservation, the opposite of von Baer's law. The same conclusion holds
6 for 1-dN and CS using *D. purpureum* expression data (both $p < 0.001$) though the
7 coefficients are not significant when using *D. discoideum* (Table 1, supplementary table
8 S2).
9

10
11
12 The lack of fit to both von Baer's law and the hourglass pattern remains when we
13 redo the above analyses using all developmental stages (supplementary table S3), or
14 using only the genes that are most specific to development (supplementary table S4, fig.
15 S1), or to individual timepoints (supplementary table S5), or to somatic (pre-stalk) tissues
16 (supplementary table S6 and fig. S2). Among these 29 new quadratic regressions there
17 was only one significant negative t^2 effect supporting an hourglass pattern, and no
18 negative effects of t in the linear regressions. Instead most of the t effects are
19 significantly positive, confirming and strengthening our initial finding of late constraint.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 **Translational Selection Does Not Support Von Baer's Law**

40
41 One further pattern reinforces the general message that temporal patterns of constraint are
42 different in *Dictyostelium* compared to animals. Gene sequence conservation is generally
43 highly correlated with expression level (Pal *et al.* 2001; Drummond *et al.* 2005; Wall *et*
44 *al.* 2005), reflecting selection for accurate translation, perhaps because of a greater
45 importance of proper protein folding in highly expressed genes (Drummond *et al.* 2005;
46 Drummond and Wilke 2008). If translational or misfolding errors early in development
47 are more serious because of cascading effects later, then we would expect sequence
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 conservation to be more related to gene expression level early in development than later
4 in development. That pattern is not found. Instead the strength of the relationship for *D.*
5
6
7
8 *discoideum* (again corrected for vegetative expression level v) peaks at intermediate
9 values (fig. 3a). This supports the hourglass pattern, but it is not a robust finding. It is
10 contradicted by strictly increasing trends when the same analysis is done using *D.*
11
12
13 *purpureum* as the focal species (fig. 3b, supplementary table S7), and when using only
14
15
16 the most developmental genes in either species (supplementary fig. S3). Again the
17
18
19 predominant pattern indicates late conservation.
20
21
22
23

24 **The timing of expression and developmental defects**

25
26
27 Given that neither the von Baer nor the hourglass patterns seems to hold, it is worth
28 asking whether the assumptions of their underlying models are met. The temporal
29
30
31 constraint model assumes that early changes have larger and therefore more deleterious
32 effects, while the spatial constraint model assumes that more deleterious changes occur at
33
34
35 some at intermediate stage. A list of genes whose mutants have known developmental
36
37
38 defects is maintained by Dictybase (<http://dictybase.org/Downloads/>). For genes with
39
40
41 stage-specific expression (2-fold higher expression at that stage than at any other), we
42
43
44 calculated the proportion that appear on the list of genes with developmental defects (fig.
45
46
47 4). This fraction was highest early in development and declined at later stages, which
48
49
50 accords best with the assumption of the temporal model.
51
52

53 **Discussion**

54
55
56
57
58
59
60

1
2
3
4 The problem of the similarity and difference of developmental stages goes back to the
5
6 early 19th century with von Baer's studies (von Baer 1828). Von Baer's law became an
7
8 important part of early evolutionary theory but thinking about the relationship between
9
10 development and evolution receded after the rise of genetics (Gould 1977). Yet the
11
12 problem of how multicellular developmental programs are constrained remains an
13
14 important one, now linked to models of network interactions and pleiotropy. A temporal
15
16 constraints model positing that early changes have pleiotropic later effects, but not vice
17
18 versa, is consistent with von Baer's law (Riedl 1978; Arthur 1988; Schank and Wimsatt
19
20 1988). A model assuming that spatial constraints are more important is consistent with
21
22 the hourglass pattern (Raff *et al.* 1991; Raff 1996), assuming that modularity late in
23
24 development reduces constraints by making most interactions local rather than global.
25
26 On the empirical side, the development of RNA-seq methods now allow us to thoroughly
27
28 catalog the genes being expressed during different developmental stages, so we can
29
30 conduct powerful studies of conservation of all genes rather than of a modest number of
31
32 morphological traits. The recent application of these techniques to plants showed
33
34 support for the hourglass pattern, even though that pattern has not been evident from
35
36 morphological studies (Quint *et al.* 2012). This raises the possibility that the hourglass
37
38 pattern might be common to many or all multicellular systems, perhaps reflecting
39
40 widespread conformity to the spatial constraints model.
41
42
43
44
45
46
47

48 We find however that *Dictyostelium* development does not follow either of the
49
50 canonical animal patterns. Instead, most analyses point to a novel pattern: greater
51
52 conservation of the genes expressed later in development. We can therefore reject both
53
54 the temporal constraint model and the standard spatial constraint model, at least as they
55
56
57
58
59
60

1
2
3 have been conceived up to this point. However the possible relationships between
4
5 patterns and process may be more complex than has been thought. The rejection of the
6
7 temporal model seems strong but the spatial constraint model, suitably modified, could
8
9 account for results. Our predominant result resembles the first half of the hourglass;
10
11 conservation increases with time but then does not show the decrease of the second half
12
13 of the hourglass. Perhaps *Dictyostelium* simply never reaches the second half, where
14
15 constraints are reduced by a break-up into relatively independent modules. Though
16
17 prespore and prestalk regions do separate relatively early, the two tissue types continue to
18
19 interact with each other, and to move with respect to each other, throughout the rest of
20
21 development. The final spatial relationship is not achieved until the culmination of
22
23 development, when the prespore cells, following signals from stalk cells in the tip, move
24
25 up to the top of the stalk (Dickinson *et al.* 2012). Final maturation of spore cells is
26
27 triggered in part by signals from stalk cells (Wang *et al.* 1999).
28
29
30
31
32
33

34 Therefore, we propose that the spatial model be divided into two: the existing spatial
35
36 model with strong modularity that predicts the hourglass pattern, and a new spatial model
37
38 for organisms with little or no modularity, which predicts a new pattern of increasing
39
40 constraint throughout development. By analogy with the hourglass pattern, von Baer's
41
42 law is sometimes called the funnel pattern (Irie and Kuratani 2011), and our new pattern
43
44 could be called the inverted funnel. Table 2 shows a classification of models, patterns,
45
46 and their relationship. Though our results could be viewed as a rejection of what has
47
48 been called the hourglass model (the spatial constraints model), they can also be viewed
49
50 as providing a generalization and strengthening of the basic logic that underlies that
51
52 model, provided the assumption of late modularity is revised. Theories are best tested not
53
54
55
56
57
58
59
60

1
2
3 by repeated confirmation of the same prediction but by novel predictions arising in
4 special cases. The most basic feature of the thinking behind the hourglass model is
5
6 neither the hourglass pattern itself nor the assumptions of a particular form of
7
8 development. Instead the main feature is the idea that the time of maximum conservation
9
10 will correspond to the time of maximum global interaction. A difference in the time of
11
12 maximum global interaction should cause a difference in the time of maximum
13
14 conservation, and that is what we appear to see in *Dictyostelium*.
15
16
17
18
19

20 Our analysis showing that genes expressed earlier in development are more likely to
21
22 have developmental defects (fig 4) weighs against this interpretation. This pattern is
23
24 more consistent with the assumptions of the temporal model. The *prima facie* predictions
25
26 for the spatial models with and without modules would be more defects for genes with
27
28 intermediate and late expression times, respectively. There may be biases in the dataset
29
30 of mutant defects, because late development has been less heavily studied (Dickinson *et*
31
32 *al.* 2012), or because severe late defects such as defective spores may be less noticeable,
33
34 but the analysis at least suggests that we should be receptive to alternative explanations.
35
36
37
38

39 As noted in the introduction, *Dictyostelium* development differs in many ways from
40
41 development in animals. The smaller soma (stalk) does not seem to be the explanation
42
43 for the failure to find the canonical patterns, because they also fail to appear when we
44
45 examine prestalk genes only (supplementary fig. S2, table S6). However, it is impossible
46
47 at this stage to rule out roles of some of the other differences in *Dictyostelium's*
48
49 development. For example, since development occurs after aggregation, the multicellular
50
51 stage can contain multiple clones, which compete to become spores instead of dead stalk.
52
53
54
55 Evolutionary conflict can lead to continual adaptive evolution and increased evolutionary
56
57
58
59
60

1
2
3 rates (Nielsen 2005) and this might explain higher evolutionary rates early in
4
5
6 *Dictyostelium* development if competition is strongest at the earlier stages when prespore
7
8 and prestalk roles are initially assigned. However, competition between clones within
9
10 fruiting bodies may be relatively infrequent, as the only study of genetic structure in the
11
12 field shows that most fruiting bodies consist of a single clone (Gilbert *et al.* 2007).
13
14

15
16 Given all the developmental differences between *Dictyostelium* and animals, it may
17
18 prove difficult to isolate low modularity as the key to its pattern of late conservation.
19
20 Gene ontology patterns can identify developmental categories of genes that are expressed
21
22 at different times (Parikh *et al.* 2010) but they do not readily distinguish global from
23
24 more local interactions. *Dictyostelium* developmental biologists should use their
25
26 expertise on the details of development to weigh in on whether the period of maximum
27
28 global interaction is early, intermediate, or late. But perhaps the best way forward would
29
30 be to test the model in other organisms with low modularity. Simple animals like corals
31
32 or *Trichoplax* might fit the model. So might simpler plants, such as fern gametophytes or
33
34 even moss sporophytes. Fungal fruiting bodies are another possibility. If the model
35
36 proves successful, it might be used as an indicator of the degree of complex modularity in
37
38 development. For example, we would expect to see the non-modular late-conservation
39
40 pattern in the simpler multicellular members of the Volvocales but, if they are sufficiently
41
42 complex, the most multicellular forms like *Volvox* would show the hourglass pattern.
43
44
45
46
47
48
49
50

51 **Materials and Methods**

52
53 **Genome and Transcriptome Data.** We retrieved predicted protein-coding gene models
54
55 for *D. discoideum* (version 03-10-2010) and *D. purpureum* (version 02-04-2010) from
56
57
58
59
60

1
2
3 dictyBase (http://dictybase.org/db/cgi-bin/dictyBase/download/blast_databases.pl).
4
5 Orthologous pairs were found using reciprocal best matches of the BLASTP using the
6
7 Inparanoid algorithm (Ostlund *et al.* 2010), as previously described (Sucgang *et al.*
8
9 2011). Parikh and colleagues provided normalized RNA-seq transcriptome data at 4-hour
10
11 intervals and from prestalk and prespore cells
12
13 (<http://dictygenome.bcm.tmc.edu/~anup/rnaseq>). For each species, we averaged the
14
15 RNA-seq read counts for each gene from the two biological replicates of each timepoint.
16
17 We eliminated genes that do not have a stage with at least 30 reads from at least one
18
19 timepoint (equal to the length of average one transcript) and genes that are not
20
21 biologically reproducible in the RNA-seq replicates. This resulted in 5259 *D. discoideum*
22
23 genes and 6014 *D. purpureum* genes with orthologs in the other species.
24
25
26
27
28
29
30
31

32 **Computational and Statistical Analysis.** We estimated nonsynonymous nucleotide
33
34 substitution rates (dN) by the maximum likelihood program *codeml* of PAML4 (Yang
35
36 2007), based on retro-translated protein sequence alignments from GAP4 (Huang and
37
38 Brutlag 2007) global alignments (synonymous changes are saturated (Sucgang *et al.*
39
40 2011)). We calculated conservation scores (CS) (Lopez-Bigas and Ouzounis 2004),
41
42 which range from 0 to 1, from GAP4 global alignments.
43
44
45

46 To estimate stage specificity of expression, we applied the tissue specificity index τ
47
48 (Yanai *et al.* 2005) to our 7 stages instead of to tissues. We normalized variables as
49
50 following: dN , square root log transform; CS and τ , arcsine square root transform; gene
51
52 length and RNA-seq read count, log transform. We used the R modules *glm* to construct
53
54
55
56
57
58
59
60

1
2
3 logistic models, and the module *lm* for linear models (R Development Core Team 2005).

4
5 We also used R modules *car* and *effects*.

6 7 8 9 10 **Supplementary Material**

11
12 Supplementary figures and tables are available at Molecular Biology and Evolution
13
14 online (<http://www.mbe.oxfordjournals.org/>).

15 16 17 18 19 20 **Acknowledgments**

21
22 We thank Gad Shaulsky for access to RNA-seq data and referees for helpful comments.

23
24 This material is based upon work supported by the National Science Foundation under
25
26 grants DEB-0918931 and DEB-1146375.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.	
	<i>D. discoideum</i>	<i>D. purpureum</i>	<i>D. discoideum</i>	<i>D. purpureum</i>
<i>O</i>	0.0050 ns ^b	-0.0060 ns	0.039 ***	0.63 ***
<i>CS</i>	0.00029 ns	0.00097 ns	0.00071 ns	0.0061 ***
<i>1-dN</i>	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.

1
2
3 **Fig. 3.** Relative translational selection at different stages. For each time point of (a) *D.*
4 *discoideum* and (b) *D. purpureum*, we regressed gene sequence conservation (*CS* or 1-
5 *dN*) on gene expression level (rescaled RNA-seq read counts) at that time, partialing out
6 the effects of level of vegetative expression level (regression data for each timepoint
7 shown in table S7). Higher translational (purifying) selection at a timepoint should be
8 reflected in a higher partial regression coefficient for that timepoint. The error bars show
9 standard errors. To show the pattern with expression time (dashed lines), we perform
10 another regression, of the partial coefficients on expression time *i* (a. *CS*, $p = 0.01$ for
11 both i^2 and *i* coefficients; 1-dN, $p = 0.001$ for both. b. i^2 terms not sig., so linear
12 regressions are shown; *CS*, $p=0.04$; 1-dN, $p=0.1$).
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 Fig. 4. The proportion of stage-specific genes associated with known developmental
30 defects. The stage-specific genes are those with at least 2-fold higher number of RNA-seq
31 reads at that stage than at any other stage, including the vegetative stage ($N = 65, 26, 105,$
32 $99, 320$). For stages 8-24, the proportion of these genes matching a list of genes with
33 known developmental defects ($N = 1,015$, <http://dictybase.org/Downloads/>, 16-03-2011)
34 is plotted.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- 1
2
3
4
5
6
7 Arthur W. 1988. A theory of the evolution of development. Chichester: John Wiley and
8
9 Sons.
10
11 Artieri CG, Haerty W, Singh RS 2009. Ontogeny and phylogeny: molecular signatures of
12
13 selection, constraint, and temporal pleiotropy in the development of *Drosophila*.
14
15 *BMC Biology* 7: 42.
16
17
18 Buss LW 1987. The evolution of individuality. Princeton: Princeton University Press.
19
20
21 Cooper TF, Ostrowski EA, Travisano M 2007. A negative relationship between mutation
22
23 pleiotropy and fitness effect in yeast. *Evolution* 61: 1495-1499. doi:
24
25 10.1111/j.1558-5646.2007.00109.x
26
27
28 Dickinson DJ, Nelson WJ, Weis WI 2012. An epithelial tissue in *Dictyostelium*
29
30 challenges the traditional origin of metazoan multicellularity. *Bioessays* 34: 833-
31
32 840.
33
34
35 Domazet-Lošo T, Tautz D 2010. A phylogenetically based transcriptome age index
36
37 mirrors ontogenetic divergence patterns. *Nature* 468: 815-819. doi:
38
39 10.1038/nature09632
40
41
42 Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH 2005. Why highly
43
44 expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* 102: 14338-14343.
45
46 doi: 10.1073/pnas.0504070102
47
48
49 Drummond DA, Wilke CO 2008. Mistranslation-induced protein misfolding as a
50
51 dominant constraint on coding-sequence evolution. *Cell* 134: 341-352. doi:
52
53 10.1016/j.cell.2008.05.042
54
55
56
57
58
59
60

- 1
2
3 Eichinger L, Pachebat JA, Glöckner G, et al. (98 co-authors) 2005. The genome of the
4
5 social amoeba *Dictyostelium discoideum*. *Nature* 435: 43-57.
6
7
8 Galis F, van Dooren TJM, Metz JAJ 2002. Conservation of the segmented germband
9
10 stage: robustness or pleiotropy. *Trends Genet.* 18: 504-509.
11
12
13 Garfield DA, Wray GA 2009. Comparative embryology without a microscope: using
14
15 genomic approaches to understand the evolution of development. *J Biol.* 8: 65.
16
17
18 Gerhart G, Kirschner M. 1997. Cells, embryos and evolution; toward a cellular and
19
20 developmental understanding of phenotypic variation and evolutionary
21
22 adaptability. Malden (MA): Blackwell Science.
23
24
25 Gilbert OM, Foster KR, Mehdiabadi NJ, Strassmann JE, Queller DC 2007. High
26
27 relatedness maintains multicellular cooperation in a social amoeba by controlling
28
29 cheater mutants. *Proc Natl Acad Sci U S A* 104: 8913-8917.
30
31
32 Gould SJ. 1977. Ontogeny and phylogeny. Cambridge (MA): Harvard University Press.
33
34
35 Grosberg RK, Strathmann RR 2007. The evolution of multicellularity: A minor major
36
37 transition? *Annu Rev Ecol Evol Syst.* 38: 621-654. doi:
38
39 10.1146/annurev.ecolsys.36.102403.114735
40
41
42 Hahn MW, Kern AD 2005. Comparative genomics of centrality and essentiality in three
43
44 eukaryotic protein-interaction networks. *Mol Biol Evol.* 22: 803-806. doi:
45
46 10.1093/molbev/msi072
47
48
49 Hazkani-Covo E, Wool D, Graur D 2005. In search of the vertebrate phylotypic stage: a
50
51 molecular examination of the developmental hourglass model and von Baer's third
52
53 law. *J Exptl Zool.* 304B: 150-158.
54
55
56
57
58
59
60

- 1
2
3 He XL, Zhang JZ 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173:
4 1885-1891. doi: 10.1534/genetics.106.060269
5
6
7
8 Huang S, Brutlag DL 2007. Dynamic use of multiple parameter sets in sequence
9 alignment. *Nucl Acids Res.* 35: 678-686.
10
11
12 Irie N, Kuratani S 2011. Comparative transcriptome analysis reveals vertebrate
13 phylotypic period during organogenesis. *Nature Commun.* 2: 248.
14
15
16
17 Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U,
18 Bergman CM, Tomancak P 2010. Gene expression divergence recapitulates the
19 developmental hourglass model. *Nature* 468: 811-814. doi: 10.1038/nature09634
20
21
22
23
24 Kato M, Shaw C, Xu QK, VanDriessche N, Morio T, Kuwayama H, Obara S,
25 Urushihara H, Tanaka Y, Shaulsky G 2004. An orderly retreat: dedifferentiation is
26 a regulated process. *Proc. Natl. Acad. Sci. U S A* 101: 7005-7010.
27
28
29
30
31
32 Kessin RH. 2001. Dictyostelium: evolution, cell biology, and the development of
33 multicellularity. Cambridge UK: Cambridge University Press.
34
35
36
37 King N 2004. The unicellular ancestry of animal development. *Devel Cell* 7: 313-325.
38
39 Lopez-Bigas N, Ouzounis CA 2004. Genome-wide identification of genes likely to be
40 involved in human genetic disease. *Nucl Acids Res.* 32: 3108-3114.
41
42
43
44 Nielsen R 2005. Molecular signatures of natural selection. *Annu Rev Genet.* 39: 197-218.
45
46 Orr HA 2000. Adaptation and the cost of complexity. *Evolution* 54: 13-20. doi:
47 10.1111/j.0014-3820.2000.tb00002.x
48
49
50
51 Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O,
52 Sonnhammer ELL 2010. InParanoid 7: new algorithms and tools for eukaryotic
53 orthology analysis. *Nucl Acids Res.* 38: D196-D203. doi: 10.1093/nar/gkp931
54
55
56
57
58
59
60

- 1
2
3 Otto S 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles.
4
5 *Proc Roy Soc B* 271: 705-714.
6
7
8 Pal C, Papp B, Hurst LD 2001. Highly expressed genes in yeast evolve slowly. *Genetics*
9
10 158: 927-931.
11
12 Parikh A, Miranda ER, Katoh-Kurasawa M, et al. (13 co-authors) 2010. Conserved
13
14 developmental transcriptomes in evolutionarily divergent species. *Genome Biology*
15
16 11: R35. doi: 10.1186/gb-2010-11-3-r35
17
18
19 Quint M, Drost H-G, Gabel A, Ullrich KK, Bönn M, Grosse I 2012. A transcriptomic
20
21 hourglass in plant embryogenesis. *Nature* 490: 98-101.
22
23
24 R Development Core Team. 2005. R: A language and environment for statistical
25
26 computing. Vienna, Austria: R Foundation for Statistical Computing.
27
28
29 Raff R, Wray G, Henry J. 1991. Implications of radical evolutionary change in early
30
31 development for concepts of developmental constraint. In: Warren L, Koprowski
32
33 H, editors. *New perspectives on evolution*. New York: Wiley-Liss. p. 189-207.
34
35
36 Raff RA. 1996. *The shape of life: genes, development, and the evolution of animal form*.
37
38 Chicago: University of Chicago Press.
39
40
41 Riedl R. 1978. *Order in living organisms: a systems analysis of evolution*. New York:
42
43 John Wiley & Sons.
44
45
46 Sander K. 1983. The evolution of patterning mechanisms: gleanings from insect
47
48 embryogenesis and spermatogenesis. In: Goodwin B, Holder N, Wylie C, editors.
49
50 *Development and evolution*. Cambridge (MA): Cambridge University Press. p.
51
52 137-159.
53
54
55
56
57
58
59
60

- 1
2
3 Schank JC, Wimsatt WC. 1988. Generative entrenchment and evolution. In: Fine A,
4
5 Machamen PK, editors. PSA-1986 Vol II: Philosophy of Science Association. p.
6
7 33-60.
8
9
10 Schlosser G, Wagner GP (editors). 2004a. Modularity in development and evolution. In.
11
12 Chicago: University of Chicago Press. p. 600.
13
14
15 Schlosser G, Wagner G. 2004b. Introduction: the modularity concept in developmental
16
17 and evolutionary biology, In: Schlosser G, Wagner GP, editors. Modularity in
18
19 development and evolution. Chicago: University of Chicago Press, pp 1-11.
20
21
22 Strassmann JE, Zhu Y, Queller DC 2000. Altruism and social cheating in the social
23
24 amoeba, *Dictyostelium discoideum*. *Nature* 408: 965-967.
25
26
27 Sugang R, Kuo A, Tian XJ, (43 co-authors) 2011. Comparative genomics of the social
28
29 amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. *Genome*
30
31 *Biology* 12: R20. doi: 10.1186/gb-2011-12-2-r20
32
33
34 Thompson CRL, Reichelt S, Kay RR 2004. A demonstration of pattern formation without
35
36 positional information in *Dictyostelium*. *Devel Growth Diff* 46: 363-369.
37
38
39 Van Dyken JD, Wade MJ 2010. The genetic signature of conditional expression.
40
41 *Genetics* 184: 557-570.
42
43
44 von Baer KE. 1828. Über Entwicklungsgeschichte der Thiere: Beobachtung und
45
46 Reflexion. Königsberg: Gebrüder Bornträger.
47
48
49 Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, Feldman MW 2005.
50
51 Functional genomic analysis of the rates of protein evolution. *Proc Natl Acad Sci*
52
53 *U S A* 102: 5483-5488. doi: 10.1073/pnas.0501761102
54
55
56
57
58
59
60

- 1
2
3 Wang N, Soderbom F, Anjard C, Shaulsky G, Loomis WF 1999. SDF-2 induction of
4
5 terminal differentiation in *Dictyostelium discoideum* is mediated by the membrane-
6
7 spanning sensor kinase DhkA. *Mol. Cell. Biol.* 19: 4750-4756.
8
9
10 Williams J. 1997. Prestalk and stalk heterogeneity in *Dictyostelium*. In: Maeda Y, Inouye
11
12 K, Takeuchi I, editors. *Dictyostelium - A model system for cell and developmental*
13
14 *biology*. Tokyo, Japan: Universal Academy Press. p. 293-304.
15
16
17 Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A,
18
19 Horn-Saban S, Safran M, Domany E, Lancet D, Shmueli O 2005. Genome-wide
20
21 midrange *transcription* profiles reveal expression level relationships in human
22
23 tissue specification. *Bioinformatics* 21: 650-659. doi:
24
25 10.1093/bioinformatics/bti042
26
27
28
29 Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular*
30
31 *Biology and Evolution* 24: 1586-1591.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

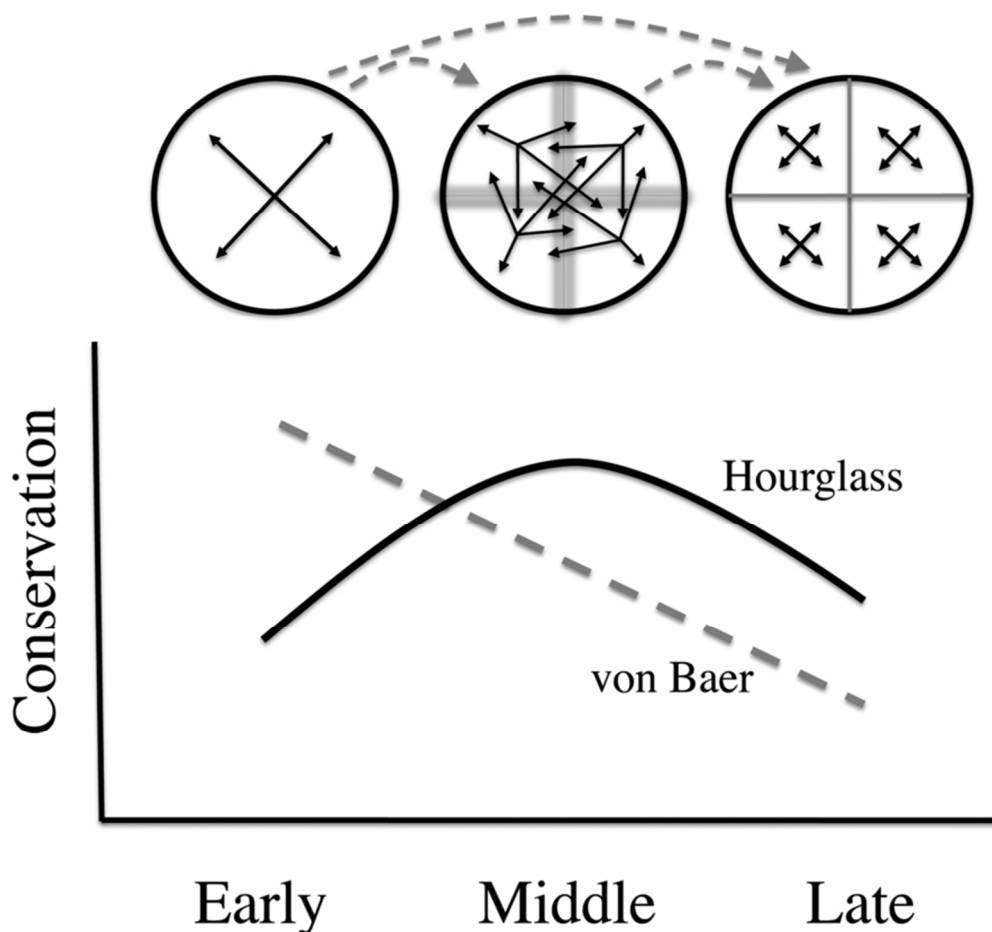


Fig. 1. Schematic of development, with proposed effects on conservation. Dashed grey arrows show that early stages affect later stages, so that early changes create large effects that are less likely to be favored by selection, leading to von Baer's law (dashed grey plot). Solid black arrows indicate 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)

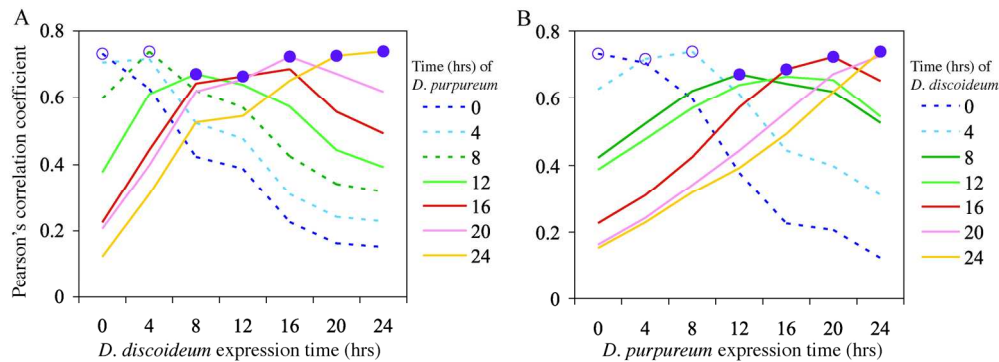


Fig. 2. Expression is more conserved later in development. Expression similarity was calculated as Pearson's correlation coefficients of normalized gene expression levels between 4-hr timepoints between the two species (Parikh et al. 2010). Each correlation is an average over the four pairwise comparisons 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 (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)

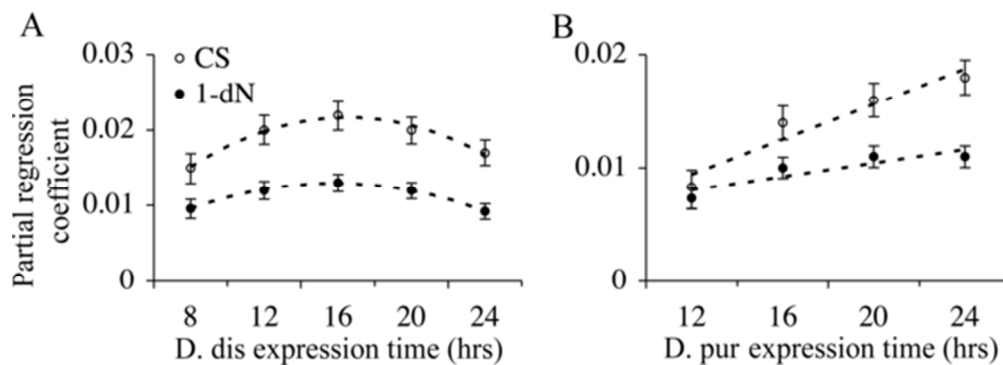


Fig. 3. Relative translational selection at different stages. For each time point of (a) *D. discoideum* and (b) *D. purpureum*, we regressed gene sequence conservation (CS or 1-dN) on gene expression level (rescaled RNA-seq read counts) at that time, partialing out the effects of level of vegetative 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)

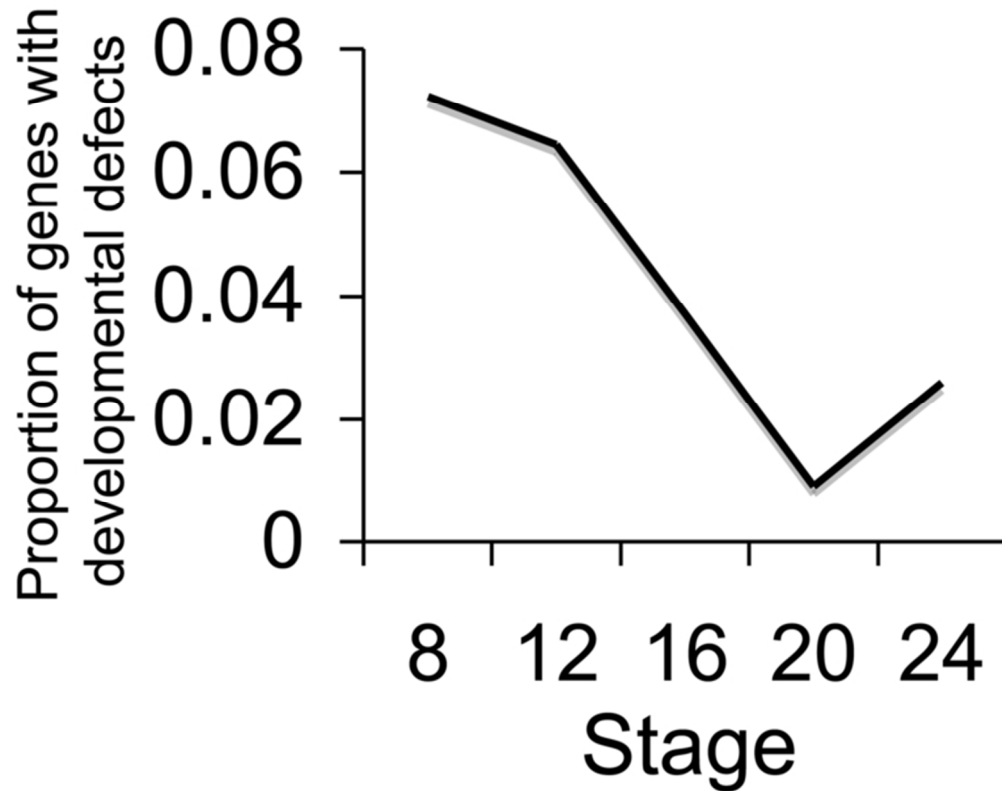


Fig. 4. The proportion of stage-specific genes associated with known developmental defects. The stage-specific genes are those with at least 2-fold higher number of RNA-seq reads at that stage than at any other stage, including the vegetative stage ($N = 65, 26, 105, 99, 320$). For stages 8-24, the proportion 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)