large collection of novel htt-interacting proteins using Y2H screens and biochemical approaches. Although it will be important to validate the interacting proteins as *bona fide* genetic modifiers of htt toxicity using different strategies and HD models, these novel interactors can provide considerable insight into the normal function of htt and the molecular pathogenesis of HD.

Acknowledgements
This work was supported by NIH grants NS36232 and NS045016.

References

1 Gusella, J.F. and MacDonald, M.E. (1995) Huntington's disease. *Semin. Cell Biol.* 6, 21–28
2 Brinkman, R.R. *et al.* (1997) The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *Am. J. Hum. Genet.* 60, 1202–1210
3 Squitieri, F. *et al.* (2001) Onset and pre-onset studies to define the Huntington's disease natural history. *Brain Res. Bull.* 56, 233–238
4 Metzger, S. *et al.* (2006) Genetic analysis of candidate genes modifying the age-at-onset Huntington's disease. *Hum. Genet.* 120, 285–292
5 Rubinsztein, D.C. *et al.* (1997) Genotypes at the GluR6 kainate receptor locus are associated with variation in the age of onset of Huntington disease. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3872–3876
6 Li, J.L. *et al.* (2003) A genome scan for modifiers of age at onset in Huntington disease: the HD MAPS study. *Am. J. Hum. Genet.* 73, 682–687
7 van Dellen, A. *et al.* (2005) Gene-environment interactions, neuronal dysfunction and pathological plasticity in Huntington's disease. *Clin. Exp. Pharmacol. Physiol.* 32, 1007–1019
8 Squitieri, F. *et al.* (2006) Juvenile Huntington's disease: does a dosage-effect pathogenic mechanism differ from the classical adult disease? *Mech. Ageing Dev.* 127, 208–212
9 Cattaneo, E. *et al.* (2005) Normal huntingtin function: an alternative approach to Huntington's disease. *Nat. Rev. Neurosci.* 6, 919–930
10 Dragatsis, I. *et al.* (2000) Inactivation of Hdh in the brain and testis results in progressive neurodegeneration and sterility in mice. *Nat. Genet.* 26, 300–306
11 Duyao, M.P. *et al.* (1995) Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science* 269, 407–410
12 Beal, M.F. and Ferrante, R.J. (2004) Experimental therapeutics in transgenic mouse models of Huntington's disease. *Nat. Rev. Neurosci.* 5, 373–384
13 Landles, C. and Bates, G.P. (2004) Huntingtin and the molecular pathogenesis of Huntington's disease. Fourth in molecular medicine review series. *EMBO Rep.* 5, 958–963
14 Harjes, P. and Wanker, E.E. (2003) The hunt for huntingtin function: interaction partners tell many different stories. *Trends Biochem. Sci.* 28, 425–433
15 Li, S.H. and Li, X.J. (2004) Huntingtin-protein interactions and the pathogenesis of Huntington's disease. *Trends Genet.* 20, 146–154
16 Borrell-Pages, M. *et al.* (2006) Huntington's disease: from huntingtin function and dysfunction to therapeutic strategies. *Cell. Mol. Life Sci.* 63, 2642–2660
17 Hyun, T.S. and Ross, T.S. (2004) HIP1: trafficking roles and regulation of tumorigenesis. *Trends Mol. Med.* 10, 194–199
18 Rong, J. *et al.* (2007) Regulation of intracellular HAP1 trafficking. *J. Neurosci. Res.* 85, 3025–3029
19 Sugars, K.L. and Rubinsztein, D.C. (2003) Transcriptional abnormalities in Huntington disease. *Trends Genet.* 19, 233–238
20 Kaltenbach, L.S. *et al.* (2007) Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genet.* 3, e82
21 Goehler, H. *et al.* (2004) A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol. Cell* 15, 853–865
22 Lim, J. *et al.* (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 125, 801–814

0168-9525/$ – see front matter
doi:10.1016/j.tig.2007.07.007

Genome Analysis

A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages

Tomislav Domazet-Lošo[1], Josip Brajković[1] and Diethard Tautz[2]

[1] Division of Molecular Biology, Ruđer Bošković Institute, Bijenička cesta 54, P.P. 180, 10002 Zagreb, Croatia
[2] Institute for Genetics, Zülpicher Straße 47, D-50674 Cologne, Germany

Macroevolutionary trends traditionally are studied by fossil analysis, comparative morphology or evo-devo approaches. With the availability of genome sequences and associated data from an increasing diversity of taxa, it is now possible to add an additional level of analysis: genomic phylostratigraphy. As an example of this approach, we use a phylogenetic framework and embryo expression data from *Drosophila* to show that grouping genes by their phylogenetic origin can uncover footprints of important adaptive events in evolution.

Introduction

Comparison of metazoan genome sequences has shown that a significant fraction of genes occurs only in defined lineages [1–8]. This implies that these genes have arisen during the evolution of the respective lineages, probably in the context of lineage specific adaptations (see Glossary). The origin of such new genes seems to occur in a punctuated manner, that is, new genes initially evolve very quickly until they become locked into a pathway [2–4]. If these genes would then retain an association with a particular pathway, one could infer their evolutionary origin on the basis of the function of the genes in extant organisms and of an assessment of their phylogenetic emergence (see Introduction in Online Supplementary Material). This is the principle of 'phylostratigraphy', which we present here as a general approach to trace evolutionary innovations using data from genome projects.

The best data for a pan-metazoan statistical evaluation of gene evolution are currently available from *Drosophila* and we have focused our analysis on this dataset. However,

Corresponding author: Domazet-Lošo, T. (tdomazet@irb.hr).
Available online 29 October 2007.

Glossary

Aptation: joint term for adaptation (i.e. a character or character complex shaped by natural selection) plus exaptation (i.e. a character or character complex that existed before it was co-opted by natural selection).
Founder genes (founders): first emergence of a gene forming the basis of a new gene lineage or gene family; the origination of founder genes might correlate with functional novelty.
Genomic phylostratigraphy: a statistical approach for reconstruction of macroevolutionary trends based on the principle of founder gene formation and punctuated emergence of protein families.
Germ layers: tissue layers in animals formed during early embryogenesis. Animals traditionally considered to have bilateral symmetry have three germ layers: ectoderm, mesoderm and endoderm.
Morpho-ontogenetic domains: distinct morphological structures that appear during the ontogeny of an organism.
Phylostratum: a set of genes from an organism that coalesce to founder genes having common phylogentic origin.
Punctuated protein family evolution: a model of genome evolution that assumes that protein families were initiated by founder genes in a scattered manner through evolutionary time.

the same approach could also be used to address various biological questions on different levels of the phylogenetic hierarchy, once reasonably complete datasets become available.

A phylostratigraphic map for *Drosophila*

The choice of phylogeny is of fundamental importance for the determination of the phylogenetic origin of the genes. The tree we use here is a trade-off between an attempt to cover the most important events in the lineage leading to *Drosophila*, reliability of phylogenetic relationships and availability of the genome data for sequence comparison (Figure S1 in Online Supplementary Material).

To assign 13 382 *Drosophila* genes to the internodes on the phylogenetic tree we used BLAST sequence similarity searches [9] against the non-redundant protein database and where necessary trace and expressed sequence tag (EST) archives (see Table S2 and Methods in Online Supplementary Material). All genes were then distributed into 12 groups according to the emergence of their founders in the phylogeny; we refer to these 12 groups as genomic phylostrata (Figure 1; Table 1). For ~30% of these genes, expression patterns during embryogenesis are known from whole mount *in situ* hybridization [10] and 2105 show regulated transcription (Table 1). Their expression patterns were annotated by a controlled vocabulary that describes the morpho-ontogenetic domains in the developing fly embryo [10]. We used the affiliation to ectoderm, endoderm and mesoderm to plot these three categories on the phylostratigraphic map (Figure 2a).

If the phylogenetic origin of the genes had no influence on the distribution of expression domains, the frequency of expression domains produced by genes in a phylostratum would be expected to approximate the frequencies in the whole sample of the fly genes (65% ectoderm, 15% endoderm, 20% mesoderm). However, in the majority of phylostrata, we detect statistically significant deviations from the expected values (Figure 2a). The most parsimonious scenario that could explain these fluctuations is that the evolutionary periods we considered were marked by different adaptive or exaptive [11] tendencies, which we uncovered because extant descendants, cumulatively analyzed in each phylostratum, might reflect ancestral functions of their founder genes [1]. Thus, although *Drosophila* embryogenesis is highly derived, it appears nonetheless to retain a significant signal with respect to the origination of these tissues.

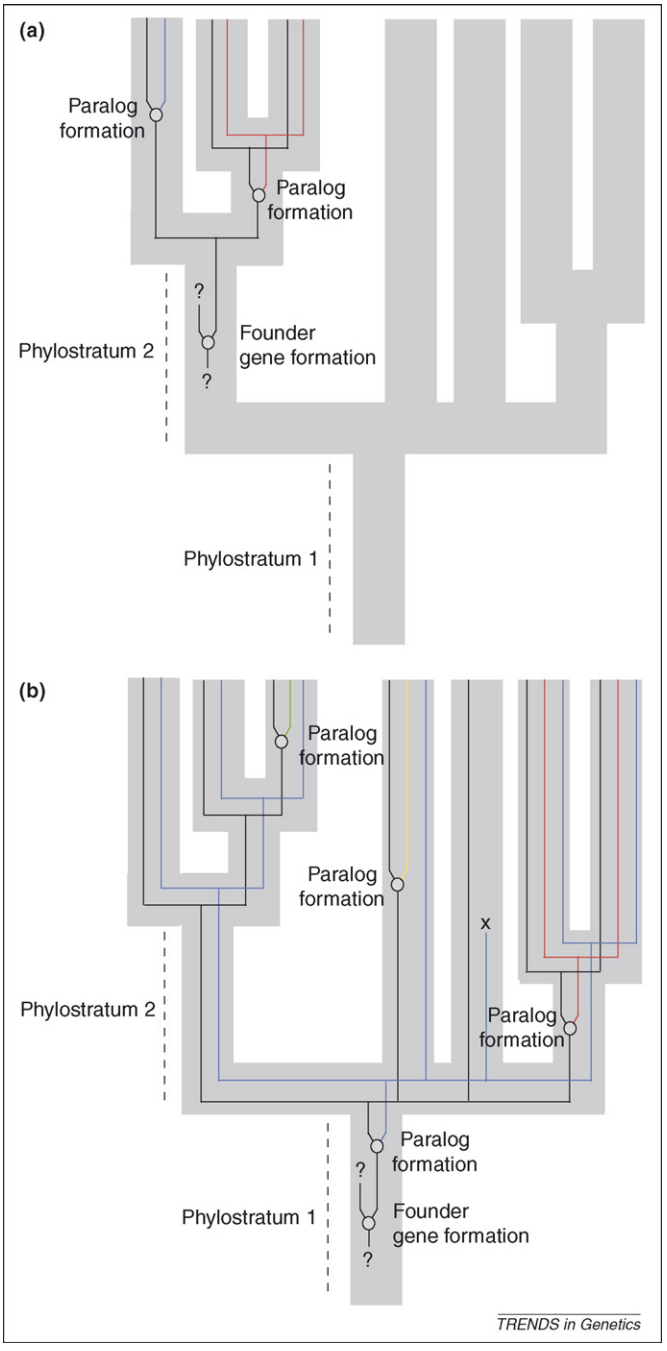

Figure 1. Model of the punctuated emergence of protein families through the formation of founder genes in different phylostrata. The gray tree represents phylogenetic relationships among hypothetical taxa, the embedded trees represent the evolution of hypothetical protein families after emergence of founder genes (see Online Supplementary Material). Circles designate the formation of the new genes, whereas different branch colors designate different paralogs. A gene loss is designated with 'X'. We show a hypothetical evolution of a gene family whose founder gene originated in the younger **(a)** and in the older internode (phylostratum) in the phylogeny **(b)**. For simplicity we show the origination of only one founder gene per phylostratum.

An ectodermal bias

The pattern of frequency fluctuations shows that on a large evolutionary scale the ectoderm appears to show

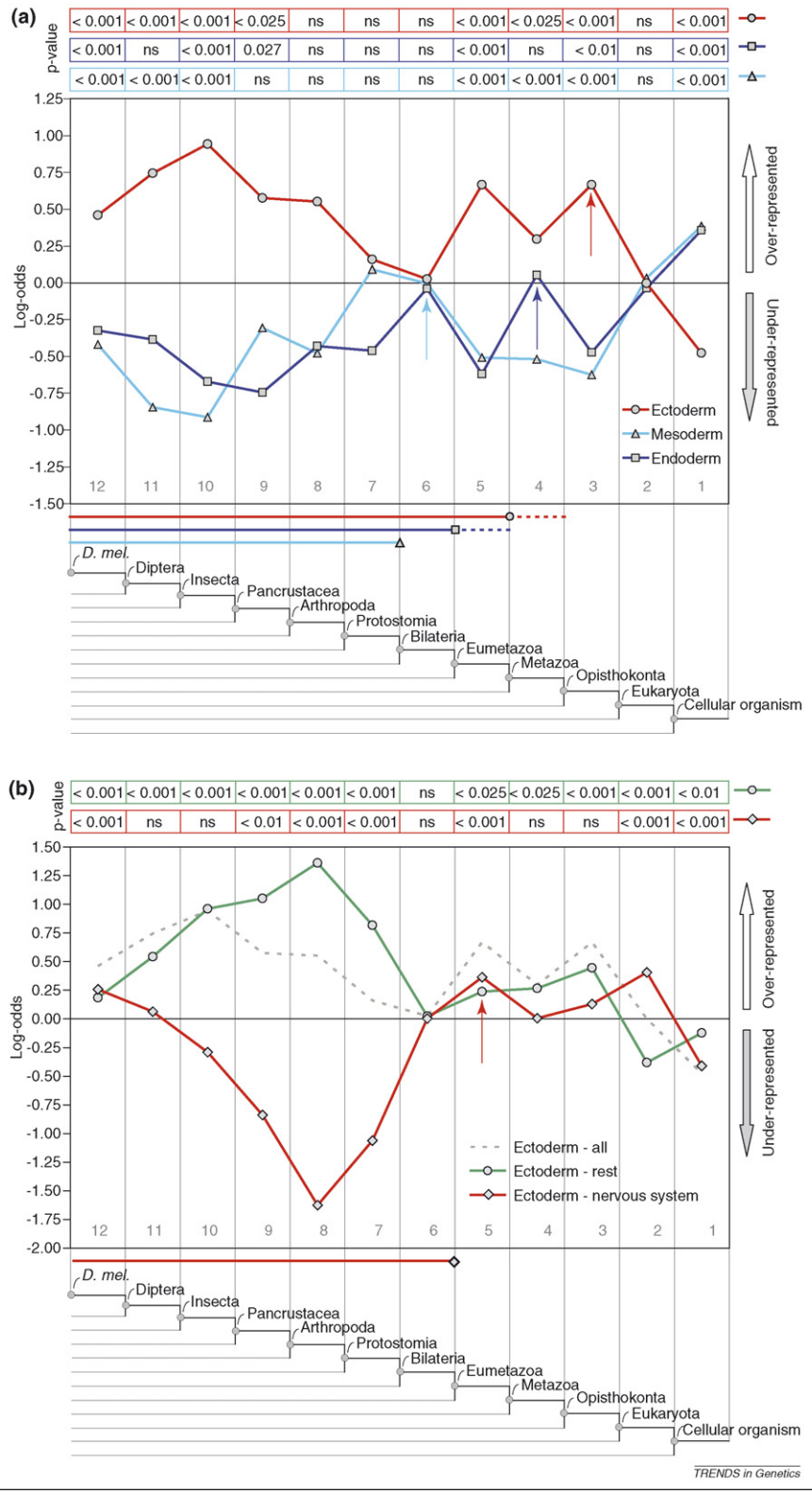

Figure 2. *Drosophila* genomic phylostratigraphic map. Twelve genomic phylostrata that correspond to the phylogenetic internodes (lower panels) are bordered by vertical grids and denote sets of *Drosophila* genes whose founder genes originated in the corresponding evolutionary periods (e.g. the leftmost phylostratum represents extant *D. melanogaster* (*D. mel.*) genes whose founders originated in the lineage leading to *D. melanogaster* after the split of dipterans, whereas the rightmost phylostratum represents genes whose founders were already present in the last common ancestor of all cellular organisms). In each phylostratum the frequency of embryo expression events is compared with the frequency in the complete sample and deviations are shown by log-odds (y-axis). Log-odds of zero denotes that the frequency of expressions in a phylostratum and in the complete sample do not differ, whereas positive and negative values point to over-representation and under-representation, respectively. Significance of the deviations is shown in the p-value chart [29] (see Methods in Online Supplementary Material). In panel **(a)**, expression frequency fluctuations for the ectoderm (red circles), endoderm (dark blue squares) and mesoderm (light blue triangles) are shown. The solid horizontal lines below the map indicate the traditional view on the first appearance of the germ layers in the evolutionary history, whereas dotted regions and colored arrows indicate distinct oscillations in frequency of expression characteristics which might be related to the origin of the germ layers. In panel **(b)**, expression frequency fluctuations for the overall ectoderm (broken, gray line), nervous system (red diamonds) and the rest of the ectoderm (green circles) are shown. The unbroken horizontal line below the map indicates the view of comparative anatomy on the origin of the nervous system. Abbreviation: ns, not significant.

Table 1. *D. melanogaster* phylostratigraphic and expression data

Phylostratum		Complete genome	Genes with *in situ* hybridization data (4141)				Germ layer analysis[a]	
			Ubiquitous expression	maternal expression	Not expressed	Restricted expression	Restricted expression	
Number	Internode	Genes (%)	Genes (%)	Genes (%)	Genes (%)	Genes (%)	Genes (%)	Expression domains
12	Diptera: ***D. melanogaster***	2356 (17.6)	12 (2.5)	31 (7.0)	244 (22.0)	156 (7.4)	142 (7.2)	930
11	Insecta: **Diptera**	467 (3.5)	7 (1.4)	7 (1.6)	47 (4.2)	68 (3.2)	61 (3.1)	303
10	Pancrustacea: **Insecta**	417 (3.1)	5 (1.0)	7 (1.6)	36 (3.3)	58 (2.8)	54 (2.8)	343
9	Arthropoda: **Pancrustacea**	78 (0.6)	1 (0.2)	0 (0.0)	5 (0.5)	22 (1.1)	21 (1.1)	102
8	Protostomia: **Arthropoda**	52(0.4)	0 (0.0)	1 (0.2)	7 (0.6)	12 (0.6)	12 (0.6)	96
7	Bilateria: **Protostomia**	134 (1.0)	0 (0.0)	0 (0.0)	18 (1.6)	22 (1.1)	21 (1.1)	148
6	Eumetazoa: **Bilateria**	1058 (7.9)	37 (7.6)	36 (8.1)	112 (10.1)	168 (8.0)	163 (8.3)	1151
5	Metazoa: **Eumetazoa**	414 (3.1)	13 (2.7)	9 (2.0)	33 (3.0)	75 (3.6)	74 (3.8)	561
4	Opisthokonta: **Metazoa**	216 (1.6)	5 (1.0)	10 (2.2)	14 (1.3)	53 (2.5)	51 (2.6)	424
3	Eukaryota: **Opisthokonta**	214 (1.6)	6 (1.2)	11 (2.5)	5 (0.5)	37(1.8)	33 (1.7)	357
2	LCA of Cellular organisms: **Eukaryota**	3105 (23.2)	205 (42.3)	154 (34.5)	193 (17.4)	536 (25.5)	498 (25.3)	4057
1	Life before LCA of cellular organisms: **LCA of Cellular organisms**	4871 (36.4)	191 (39.4)	180 (40.4)	394 (35.6)	898 (42.7)	837 (42.6)	5560
	Total	**13382 (100)**	**482 (100)**	**446 (100)**	**1108 (100)**	**2105 (100)**	**1967**	**14032**

[a]Fraction of genes with restricted expression (93%).

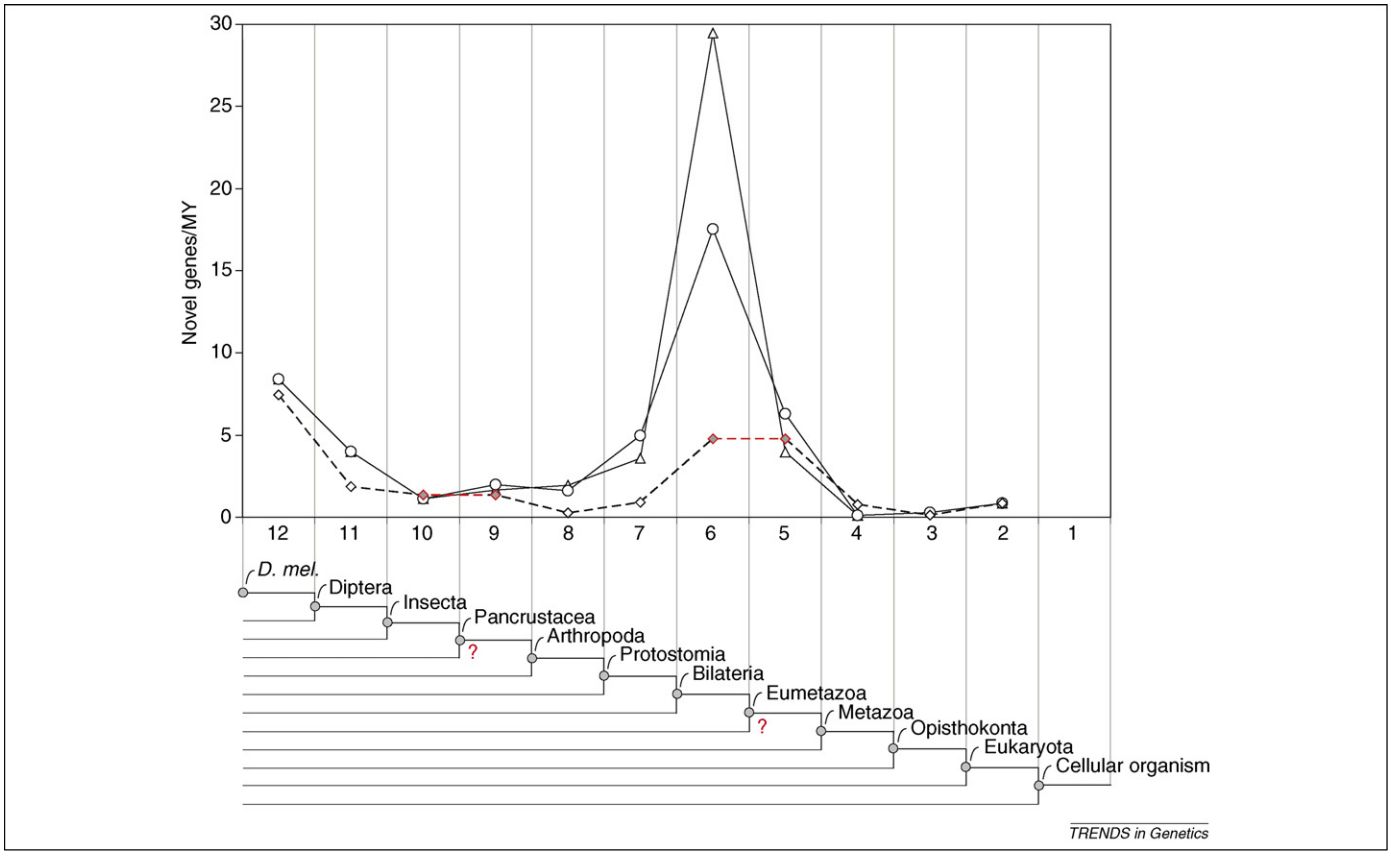

Figure 3. Fixation rates of founder genes estimated in the *D. melangaster* (*D. mel.*) genomic phylostrata. We estimated the rate of fixation of founder genes using the molecular clock time estimates from two studies (circles [18], triangles [17]), which cover neighboring nodes around phylostratum 6, and one conservative study [19] with incomplete time estimates (red question marks represent missing time estimates in this case, whereas red line regions represent approximations obtained by averaging over the two phylostrata). Time estimates for the other nodes were compiled from several sources (see Table S2 in Online Supplementary Material). X-axis numbering denote phylostrata.

an opposite trend to the endoderm and the mesoderm. When a gene coalesces to a founder gene that was generated after the diversification of the eukaryotes (phylostrata 3 to 12), there is a greater chance that it is expressed in ectodermal structures. By contrast, coalescence to a founder gene that was already present before diversification of the cellular organisms (phylostratum 1) correlates with mesodermal and endodermal expression. This pattern suggests that aptations [11], achieved by founder genes, might have been biased towards ectoderm through most of animal evolutionary history.

When one differentiates the ectodermal genes into those that are expressed in the nervous system versus the rest, one finds an additional pattern for the nervous system genes (Figure 2b). For these, a first peak is seen in phylostratum 2, representing the last common ancestor (LCA) of eukaryotes and a second one in phylostratum 5 at the time of emergence of the eumetazoa. The presence of the first peak would suggest that genes relevant for the nervous system became already available before complex nervous systems evolved. This is a rather unexpected finding that deserves further analysis in the future. By contrast, the second peak is at an expected position, namely before the emergence of the bilateria, which might well have been preceded by an elaboration of the nervous system. Interestingly, there is a strong underrepresentation of nervous system genes in phylostrata 7 to 10. This could indicate that no major genomic novelties were generated during this phase, which represents essentially the evolution of arthropods in the aquatic environment, before colonizing the land.

The last common bilaterian ancestor

A special situation exists for phylostratum 6, the time of the bilaterian LCA. Here, the genes have unbiased distribution of expression characteristics among all three germ layers (Figure 2). This finding, as well as an increased percentage (~8%) of founder genes in this period, suggests that the shift from the eumetazoan ancestor to the organizational level of the bilaterian LCA required a substantial amount of gene innovations, with balanced action among germ layers. Therefore, before rewiring of genetic modules took place during the Cambrian explosion [12–16] some other trigger might have filled up the arsenal of available genes.

Furthermore, although dating by molecular clocks should be treated with caution, we detected, using the time estimates from the studies that covered the neighboring nodes [17–19], increased fixation rate of founder genes in phylostratum 6 (Figure 3; Table S1 in Online Supplementary Material). An adequate supply of duplicated genes and suitable selective conditions might have been prerequisites for such high fixation rates of the founder genes. Complete genome duplication as well as single gene duplications [8] could provide enough raw material for selection to act on, whereas strong and varying selective pressures are easy to imagine in the unstable Neoproterozoic environment (1000 to 542 million years ago)

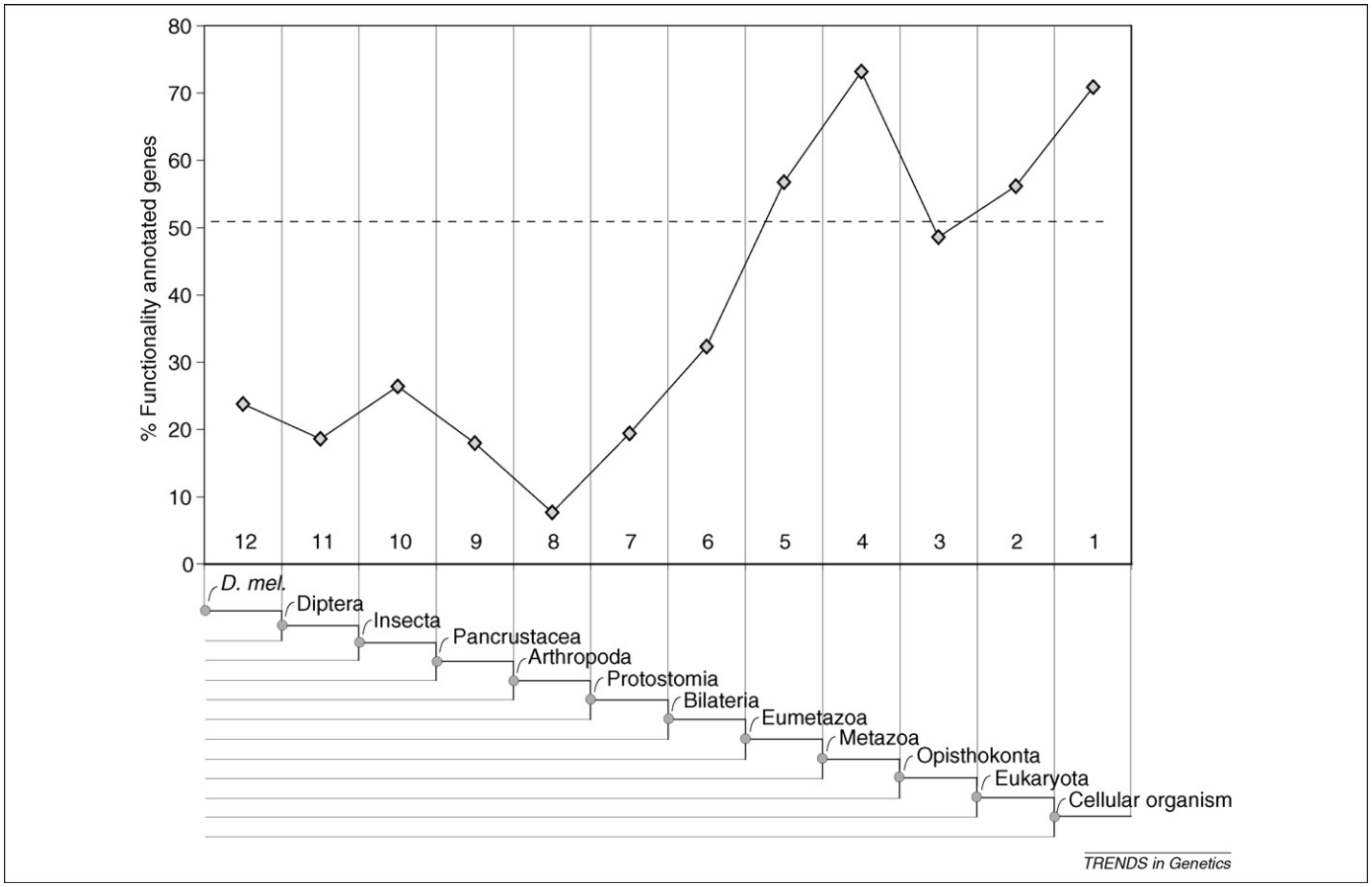

Figure 4. Difference in functional annotation of *D. melanogaster* (*D. mel.*) genes among phylostrata. The broken line depicts the average fraction of fly genes in the complete genome, which is functionally annotated with Gene Ontology biological process terms.

[13,14,16]. Interestingly, fly genes that coalesce to the founders in phylostratum 6 are considerably less functionally characterized than the older genes (Figure 4), which indicates that for a better understanding of the protostome–deuterostome ancestor, of the Cambrian explosion and of the emergence of the insects, functional studies should focus on these genes.

Origin of the germ layers

The scattered origin of founder genes that we found here suggests that the emergence of morphological innovations is coupled not only with the rewiring of existing genetic elements [12] but also with the formation of founder genes that have a specific role in the emerging gene networks. As the origin of the germ layers is a long-standing issue, it is interesting to look for possible signatures of the origin of the germ layers that might be detectable on the phylostratigraphic map in the form of a peak or an abrupt increase in frequency of expression characteristics. Indeed, we find unique and significant fluctuations that could be related to the origin of the germ layers and, interestingly, they are arranged in ectoderm-endoderm-mesoderm stepwise progression (Figure 2a). Specifically, the ectoderm in phylostratum 3 rapidly reaches an overrepresentation plateau; endoderm has a peak in phylostratum 4; and finally mesoderm has a distinct peak which spreads over phylostrata 6 and 7 (Figure 2a).

Recent studies suggest that multicellularity evolved independently in animals and fungi from unicellular opisthokont ancestors [20]. However, the presence of proteins essential for animal multicellularity in unicellular opisthokonts [21], their facultative multicellular behavior [22–24], parasitism of some lineages [20] and multicellularity of earlier branching groups [21] might challenge this view. Indeed, our finding that a higher than expected number of proteins that play a role in the ectoderm emerged in the opisthokont LCA, suggests that the opisthokont LCA displayed some sort of multicellularity. Similarly, our finding of a distinct endodermal peak is consistent with recent developmental studies in sponges, which suggests an early origin of gastrulation [25,26].

For a better understanding of the origin of the mesoderm, a more reliable phylogenetic positioning and genome sequence of Ctenophora is required [27]. However, given that cnidarian genomic data are included in the analysis, the phylostratigraphic mesodermal peak in phylostratum 6 supports the idea that mesoderm is a genuine bilaterian innovation [27,28]. These studies also showed that some mesodermal genes existed before the mesoderm emerged [27,28], which demonstrates that phylostratigraphy, as a statistical approach, is robust to functional changes of individual genes.

Concluding remarks

The finding of statistically significant differences in gene emergence in the phylostratigraphic map validates the assumption that at least a proportion of these genes have retained a signal of their evolutionary history. Therefore, under the assumption of a generality of the underlying principle of founder gene formation and punctuated emergence of protein families, genomic phylostratigraphy might also uncover macroevolutionary processes in other lineages. Evidently, the analysis and interpretations presented here are necessary preliminarily, because we relied only on the *Drosophila* data and limited phylogenetic resolution of the phylostrata. Therefore, our current findings are an example of how such an approach might uncover macroevolutionary processes in lineages once more genomic and other high-throughput data are available. The principle of genomic phylostratigraphy implies that the genome of every extant species retains parts of the picture of the evolutionary epochs. Hence, the rapid accumulation of genomic data on a large variety of organisms will allow an increasing view of these pictures.

Acknowledgements

We thank Đurđica Ugarković for support, Krunoslav Brčić-Kostić for helpful insights and discussions, Mirjana Domazet-Lošo for help with the database queries, Marko Madunić for providing literature, Nikola Pavković for help with cluster and grid framework, Bojan Basrak for statistical advice, Jonathan Howard, Treasa Creavin and four anonymous reviewers for comments that improved the manuscript. Computational resources were provided by CroGrid project (RBI), Isabella cluster (University computing center – SRCE) and KONČAR - Electronics and Informatics Inc. This study was supported by Research Fund of Republic of Croatia 098–0982913–2832 grant to Đ. Ugarković.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.08.014.

References

1 Altenberg, L. (1995) Genome growth and the evolution of the genotype-phenotype map. In *Evolution and Biocomputation: Computational Models of Evolution, LNCS* (Vol. 899) Banzhaf, W. and Eeckman, F.H.,eds In pp. 205–259, Springer Verlag
2 Domazet-Lošo, T. and Tautz, D. (2003) An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res.* 13, 2213–2219
3 Long, M. *et al.* (2003) The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* 4, 865–875
4 Choi, I.G. and Kim, S.H. (2006) Evolution of protein structural classes and protein sequence families. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14056–14061
5 Force, A. *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
6 Thornton, J.W. (2005) New genes, new functions: Gene family evolution and phylogenetics. In *Evolutionary Genetics: Concepts and Case Studies* (Fox, C. and Wolf, J., eds), pp. 157–172, Oxford University Press
7 Prince, V.E. and Pickett, F.B. (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* 3, 827–837
8 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
9 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
10 Tomancak, P. *et al.* (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* 3, research0088.1-0088.14
11 Gould, S.J. and Vrba, E.S. (1982) Exaptation - a missing term in the science of form. *Paleobiology* 8, 4–15
12 Carroll, S.B. (2005) Evolution at two levels: On genes and form. *PLoS Biol.* 3, e245
13 Knoll, A.H. and Carroll, S.B. (1999) Early animal evolution: Emerging views from comparative biology and geology. *Science* 284, 2129–2137
14 Erwin, D.H. and Davidson, E.H. (2002) The last common bilaterian ancestor. *Development* 129, 3021–3032

15 Davidson, E.H. and Erwin, D.H. (2006) Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800
16 Marshall, C.R. (2006) Explaining the Cambrian "explosion" of animals. *Annu. Rev. Earth Planet. Sci.* 34, 355–384
17 Peterson, K.J. and Butterfield, N.J. (2005) Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9547–9552
18 Peterson, K.J. *et al.* (2004) Estimating metazoan divergence times with a molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6536–6541
19 Douzery, E.J.P. *et al.* (2004) The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. U. S. A.* 101, 15386–15391
20 Steenkamp, E.T. *et al.* (2006) The protistan origins of animals and fungi. *Mol. Biol. Evol.* 23, 93–106
21 King, N. (2004) The unicellular ancestry of animal development. *Dev. Cell* 7, 313–325
22 Leadbeater, B.S.C. (1983) Life-History and ultrastructure of a new marine species of *Proterospongia* (Choanoflagellida). *J. Mar. Biol. Ass. U. K.* 63, 135–160
23 Hibberd, D.J. (1975) Observations on the ultrastructure of the choanoflagellate *Codosiga botrytis* (Ehr.) Saville-Kent with special reference to the flagellar apparatus. *J. Cell Sci.* 17, 191–219
24 Whisler, H.C. (1968) Developmental control of *Amoebidium parasiticum*. *Dev. Biol.* 17, 562–570
25 Leys, S.P. and Ereskovsky, A.V. (2006) Embryogenesis and larval differentiation in sponges. *Can. J. Zool.* 84, 262–287
26 Leys, S.P. and Eerkes-Medrano, D. (2005) Gastrulation in calcareous sponges: In search of Haeckel's gastraea. *Integr. Comp. Biol.* 45, 342–351
27 Martindale, M.Q. (2005) The evolution of metazoan axial properties. *Nat. Rev. Genet.* 6, 917–927
28 Martindale, M.Q. *et al.* (2004) Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone *Nematostella vectensis* (phylum, Cnidaria; class, Anthozoa). *Development* 131, 2463–2474
29 Castillo-Davis, C.I. and Hartl, D.L. (2003) GeneMerge–post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics* 19, 891–892

0168-9525/$ – see front matter
doi:10.1016/j.tig.2007.08.014

Crossover interference underlies sex differences in recombination rates

Petko M. Petkov[1], Karl W. Broman[2], Jin P. Szatkiewicz[1] and Kenneth Paigen[1]

[1] Center for Genome Dynamics, The Jackson Laboratory, Bar Harbor, ME 04609, USA
[2] Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205, USA

In many organisms, recombination rates differ between the two sexes. Here we show that in mice, this is because of a shorter genomic interference distance in females than in males, measured in Mb. However, the interference distance is the same in terms of bivalent length. We propose a model in which the interference distance in the two sexes reflects the compaction of chromosomes at the pachytene stage of meiosis.

Introduction

Meiosis consists of two consecutive cell divisions after a single round of DNA replication, thereby ensuring reduction of the chromosome number to produce haploid gametes. This reduction occurs in the first meiotic division, when homologous chromosomes are joined together in prophase to form bivalents and eventually separate in anaphase. In mammals, higher plants and yeast, chromosome recognition and formation of the synaptonemal complex is initiated by double-strand breaks on one chromatid. These breaks are repaired by homologous recombination, leading to genetic crossing over and/or gene conversion when a non-sister chromatid is used as a template. Given the segregation of chromatids into haploid gametes, only half of the genetically recombinant chromosomes that result from molecular recombination events will be detected.

Crossover events are not randomly spaced along chromosomes. Instead, the presence of one crossover event on a chromosome reduces the possibility of a second event nearby [1–3], a phenomenon known as crossover interference. In many species, recombination rates differ in the two sexes. The female recombination map is 1.7 times longer than that of males in humans [4,5] and 1.3 times longer in mice [6]. Several mechanisms have been proposed to play important roles: haploid selection [7]; different epistatic interactions among genes expressed during male and female meiosis [8]; presence of X-linked modifiers [9]; and regional differences in the chromatin structure of male and female gametocytes [10]. However, experimental evidence in support of these suggestions has remained elusive. Here we show that crossover interference in meiosis is the main factor underlying sex differences of recombination rates, and that the average intercrossover distance is the same in both sexes when measured in micrometers of synaptonemal complex length.

Distribution of recombination events along mouse chromosome 1

Recombination rates in each sex were measured in backcrosses of C57BL/6JxCAST/EiJ F1 male and female mice to C57BL/6J. The entirety of mouse chromosome 1 (Chr 1) was examined at ~7 Mb resolution, which ensured the detection of virtually all crossovers taking into account the strong positive interference in mouse recombination [11]. In total, we detected 2715 recombination events in 2762 progeny of female F1 parents and 1509 recombination events in 1881 progeny of male F1 parents. The average recombination rates were 0.51 cM per Mb (cM/Mb) in

Corresponding author: Petkov, P.M. (Petko.Petkov@jax.org).
Available online 26 October 2007.