# Maternal Expression Relaxes Constraint on Innovation of the Anterior Determinant, *bicoid*

Michael S. Barker[⊚], Jeffery P. Demuth[⊚*], Michael J. Wade[⊚]

Department of Biology, Indiana University, Bloomington, Indiana, United States of America

**The origin of evolutionary novelty is believed to involve both positive selection and relaxed developmental constraint. In flies, the redesign of anterior patterning during embryogenesis is a major developmental innovation and the rapidly evolving *Hox* gene, *bicoid (bcd)*, plays a critical role. We report evidence for relaxation of selective constraint acting on *bicoid* as a result of its maternal pattern of gene expression. Evolutionary theory predicts 2-fold greater sequence diversity for maternal effect genes than for zygotically expressed genes, because natural selection is only half as effective acting on autosomal genes expressed in one sex as it is on genes expressed in both sexes. We sample an individual from ten populations of *Drosophila melanogaster* and nine populations of *D. simulans* for polymorphism in the tandem gene duplicates *bcd*, which is maternally expressed, and *zerknüllt (zen)*, which is zygotically expressed. In both species, we find the ratio of *bcd* to *zen* nucleotide diversity to be two or more in the coding regions but one in the noncoding regions, providing the first quantitative support for the theoretical prediction of relaxed selective constraint on maternal-effect genes resulting from sex-limited expression. Our results suggest that the accelerated rate of evolution observed for *bcd* is owing, at least partly, to variation generated by relaxed selective constraint.**

## Introduction

The origin of novel structures and functions has long been problematic for evolutionary biology. A century-long debate has centered on the relative contribution of deterministic processes, such as natural selection, and stochastic processes, such as mutation and random genetic drift [1,2] to biological innovation. In modern parlance, "There are problems specific for the origin of novelties that are not the same as in the case of adaptive modification of existing structures ... novelties apparently arise in spite of strong developmental constraints that generally canalize morphological evolution" [3]. How descendant lineages overcome ancestral constraint is a principal research objective in the study of evolutionary novelty.

One route to gaining novel structures and functions is by relaxing ancestral evolutionary constraints. Although the degree of evolutionary lability can be discerned from the phylogenetic distribution of a given trait, the inference of relaxed constraint is more difficult and potentially tauto-logical [4,5]. We show how to test the hypothesis of relaxed constraint using population genetic theory and sequence data from the novel *Hox* gene, *bicoid (bcd),* a critical component of the redesign of anterior patterning during fly embryogenesis, a major developmental innovation [6,7].

The gene *bcd* is present only in the derived Cyclorrhaphan flies, and is the result of duplication and divergence of a more widely conserved *Hox3* descendant *zerknüllt (zen)* [7,8]. After the origin of myriapods in the arthropod lineage leading to insects, *Hox3* underwent a functional transformation to what is now called *zen* [9,10]. In taxa prior to the duplication giving rise to *bcd, zen* is expressed both maternally and zygotically and is important for early designation of both embryonic and extraembryonic tissues [11]. Following duplication, the ancestral maternal-zygotic *zen* expression pattern was partitioned into solely maternally expressed *bcd* and solely zygotic expressed *zen* [6,8]. Since the partitioning of gene expression, *bcd* has evolved faster than *zen* [12,13].

The preservation of both gene copies and the partitioning of the ancestral maternal-zygotic expression pattern between them suggest that functional constraints on both genes remain despite *bcd* evolving additional, unique functionality. Furthermore, because both are homeotic genes critical to early development, selection on the coding regions of these genes is likely to be predominantly purifying (i.e., the selection coefficient, *s,* has tended to be negative). Thus, we can reasonably assume that the effective population size, $N_e$, the regional chromosomal environment (affecting mutation and recombination), and the average magnitude and sign of *s* have been approximately the same for both genes. There are several factors that might contribute to violations of our assumption of equivalent *s*. However, it is our purpose to test whether the reduced efficacy of selection that attends sex-limited expression is sufficient to explain the pattern of *bcd* diversity without invoking changes to *s*.

From a population genetics perspective, evolutionary constraint can be relaxed by either reducing the strength of

## Synopsis

How do novel structures and functions originate? This question has proven more difficult to answer than the question of how existing structures are refined to better suit the environment. Evolution by natural selection explains the latter. Ironically, it is the power of natural selection to maintain genes that are suitable for their current roles that also works against the evolution of entirely new traits. The conservation of genes controlling the early stages of embryo development, from worms, to flies, to humans, is a famous example of natural selection's power to constrain evolution and thwart the spread of evolutionary novelties. However, the gene determining which end of the fly embryo becomes the head has a relatively recent origin. How did this gene, *bicoid,* escape purifying selection and take on a novel function? The authors investigate the hypothesis that because *bicoid* is expressed only in females it experiences only half as much constraint as a gene expressed in both sexes. Comparing sequences of *bicoid* with its duplicate gene *zerknüllt,* which retains expression in both sexes, the authors show that, as expected, the variation in *bicoid* is twice that of *zerknüllt.* The findings suggest that relaxed constraint is an important step in the origin of novel function.

purifying selection or limiting the proportion of genes exposed to selection. Here we refer to the former as relaxed selection and the latter as relaxed constraint. Population genetic theory shows that deleterious alleles at autosomal loci with sex-limited effects have equilibrium frequencies twice as high as those expressed in both sexes with similar fitness effects [14,15]. This relaxed constraint is due to deleterious alleles being "hidden" from selection in half of the population (i.e., males, assuming 1:1 sex ratio) in much the same way that recessive alleles have higher expected equilibrium frequencies (i.e., relaxed constraint) because they are "masked" from selection in heterozygotes. If *bcd* is experiencing relaxed constraint because it has sex-limited expression, then we expect the coding regions of *bcd* to be twice as diverse as those of *zen* [13]. Here, we report the first, to our knowledge, test of the prediction that maternal-effect genes have a 2-fold higher heterozygosity at mutation-selection balance than zygotically expressed genes by comparing within species sequence diversity for *bcd* and *zen* from ten populations of *Drosophila melanogaster* and nine of *D. simulans.*

## Results

At mutation-drift balance, the expected value of Tajima's $D$ is near zero [16,17] with positive (balancing selection) or negative (purifying selection) values caused by selection. The largest deviation from zero among our estimates of Tajima's $D$ is −1.01513 for *D. melanogaster bcd,* and the smallest is −0.42264 for *D. melanogaster zen.* None of the values are significantly different from zero (or from one another), suggesting both genes in both species are near equilibrium (Table 1).

The observed *bcd/zen* ratios of total sequence diversity are 1.82 for *D. melanogaster* and 1.60 for *D. simulans.* For the noncoding regions, the ratios are 0.74 and 1.28, respectively, while coding diversity ratios are 1.67 and 1.90 (Table 1). Further partitioning the diversity in coding regions, the *bcd/zen* diversity ratios for synonymous sites are 1.28 (0.00279/0.00218) for *D. melanogaster* and 1.35 (0.0229/0.017) for *D. simulans.* For

the nonsynonymous sites in coding regions, the ratios are 2.18 (0.00107/0.00049) and 2.68 (0.00918/0.00342), respectively. Combining the nucleotide diversity of noncoding and synonymous sites for each gene in each species we find the *bcd/zen* ratios at silent sites to be 1.08 for *D. melanogaster* and 0.89 for *D. simulans,* very close to the theoretical expectation of 1.0. In contrast, for nonsynonymous sites in the coding regions, the ratios are somewhat in excess of 2, the theoretically predicted value.

## Discussion

The redesign of anterior patterning during embryogenesis is a major developmental innovation, involving both positive selection and the overcoming of developmental constraints. A previous study of sequence variation in *bcd* from a Zimbabwe population of *D. melanogaster* [18] found a significant excess of replacement polymorphisms, indicative of reduced purifying selection, as well as "extensive" haplotype structure consistent with positive selection and linkage disequilibrium. The authors concluded that the selective history of *bcd* is "Complex, involving both relaxation of purifying selection in some parts of the protein and the action of selection in other parts of the gene region" [18]. We do not find evidence of haplotype structure; however, this absence is likely the result of sampling single flies from geographically diverse populations instead of the more intensive sampling of a single population, as in the study of Zimbabwe *D. melanogaster.*

Our comparison of *bcd* and *zen* sequence diversity suggests that the variation necessary for *bcd* to overcome developmental constraint is not due solely to redundancy attending gene duplication [13], but is more consistent with the relaxed constraint that accompanies sex-limited expression. Furthermore, our results indicate that most of the difference in *bcd* and *zen* variation can be accounted for by relaxed constraint without invoking differences in *s.* Although we find no extensive haplotype structure in our analysis, its existence in prior studies may be due to occasional selective sweeps that fix the high standing diversity of *bcd* within populations.

In the origin of evolutionary novelty, the balance between positive selection and relaxed constraint may be different for maternally expressed genes than for paternally expressed genes [19,20]. Although similarly expressed in only half of the members of a population, sexual selection acting on males through paternally expressed genes can be much stronger than the viability selection acting on embryos as a result of maternal-effect genes [21]. Maternal-effect genes are more similar in their evolutionary genetics to aging genes expressed late in life. Alleles with detrimental effects late in life accumulate in a population because selection against them is weaker than it is for alleles with deleterious effects of similar magnitude expressed earlier [22,23]. However, it is difficult to make a quantitative prediction regarding the elevated heterozygosity of aging genes because the degree of relaxed selective constraint depends upon a host of demographic variables. For maternal effect genes, we expect a 2-fold increase in heterozygosity at mutation-selection balance, because they are "hidden" from natural selection in half the population, i.e., in males, where they are not expressed. The levels of *bcd* and *zen* sequence diversity we observed in *D. melanogaster* and *D. simulans* support this theoretical prediction.

**Table 1.** Tajima's $D$ and Nucleotide Diversities ($\pi$) for *bcd* and *zen*

| Species | Gene | Tajima's $D$ [95% CI] | $\pi_{Total}$ | Coding Regions [a] | | | $\pi_{nc}$ | $\pi_{silent}$ |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | $\pi_s$ | $\pi_a$ | $\pi_{coding}$ | | |
| *D. melanogaster* | *bcd* | −1.01513 [−1.71, 1.83] | 0.00333 (34; 3,129) | 0.00279 (4; 342) | 0.00107 (4; 1,125) | 0.00147 (8; 1,467) | 0.00497 (26; 1,662) | 0.00460 (30; 2,004) |
| | *zen* | −0.27951 [−1.67, 1.76] | 0.00183 (7; 1,263) | 0.00218 (1; 244) | 0.00049 (2; 815) | 0.00088 (3; 1,059) | 0.00693 (4; 204) | 0.00426 (5; 448) |
| | *bcd/zen* | | 1.82 | 1.28 | 2.18 | 1.67 | 0.74 | 1.08 |
| *D. simulans* | *bcd* | −0.41719 [−1.65, 1.53] | 0.01069 (113; 3,598) | 0.0229 (24; 339) | 0.00918 (36; 1,116) | 0.0124 (60; 1,455) | 0.00955 (53; 2,117) | 0.01139 (77; 2,456) |
| | *zen* | 0.42264 [−1.7, 1.67] | 0.00669 (21; 1,254) | 0.017 (11; 241) | 0.00342 (7; 812) | 0.00654 (18; 1,053) | 0.00746 (3; 201) | 0.01269 (14; 442) |
| | *bcd/zen* | | 1.60 | 1.35 | 2.68 | 1.90 | 1.28 | 0.89 |

$\pi_a$, non-synonymous nucleotide diversity; $\pi_{coding}$, combined synonymous and nonsynonymous diversity; $\pi_{nc}$, noncoding diversity; $\pi_s$, synonymous nucleotide diversity; $\pi_{Silent}$, combined synonymous and noncoding diversity; and $\pi_{Total}$, nucleotide diversity of the complete sequence.
[a] Numbers in parentheses indicate (number of substitutions; total number of sites)
DOI: 10.1371/journal.pgen.0010057.t001

## Materials and Methods

We purchased ten *D. melanogaster* and nine *D. simulans* wild-type strains from the Tucson *Drosophila* Species Stock Center (http://stockcenter.arl.arizona.edu/) (Table S1). We selected strains from a broad geographic range to sample genetic variation across the distribution of each species. From each strain, a single individual was randomly selected for DNA extraction and analyses of *bcd* and *zen* sequence diversity.

Genomic DNA was extracted from single flies using the Puregene tissue kit (Gentra Systems, Minneapolis, Minnesota, United States). Oligonucleotides for PCR amplification and direct sequencing were designed using the FastPCR program [24] from previously published *D. melanogaster* and *D. simulans bcd* and *zen* sequences. To increase primer specificity and avoid amplifying other genes, candidate primers were searched against the *D. melanogaster* genome to remove primers that may amplify other genes. Further, in silico PCR simulations of candidate primer pairs against all members of the Antennapedia Complex were conducted in FastPCR [24] to remove primers that may amplify related sequences.

The *bcd* primer pairs amplified a ∼4-kb region that included ∼60 bp of the 5′ and ∼1 kb of the 3′ flanking regions. The *zen* primer pairs amplified an ∼1.3-kb region that included ∼30 bp of the 5′ and ∼100 bp of the 3′ flanking regions. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, California, United States). Forward and reverse strands were sequenced on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, California, United States) using Big Dye Terminator v 3.1 chemistry (Applied Biosystems). Oligonucleotides used for PCR amplifications were also used for sequencing the *D. melanogaster* and *D. simulans bcd* and *zen* genes. The resulting sequences are deposited in GenBank (http://www.ncbi.nlm.nih.gov/) with the following accession numbers: *D. melanogaster bcd* (DQ114428–DQ114437), *D. melanogaster zen* (DQ114438–DQ114447), *D. simulans bcd* (DQ114448–DQ114456), and *D. simulans zen* (DQ114457–DQ114465).

We constructed contigs for each gene using CodonCode Aligner v. 1.3.4 (CodonCode, Dedham, Massachusetts, United States). All base calls in each contig were based upon at least one forward and reverse sequence pair, with most bases covered by two forward and reverse pairs. Contigs of each gene were aligned using BioEdit v. 7.0.4.1 [25]. We verified each alignment and variable sites manually. Coding regions were defined for *bcd* following Baines et al. [18]; i.e., only the major transcript was considered coding. For our *bcd* sequences, we defined the coding regions of *D. melanogaster* as positions 68–232, 792–867, 923–1,864, and 2,377–2,663. For our *D. simulans bcd,* the coding regions were defined as positions 68–232, 776–851, 907–1,848, and 2,235–2,521. Coding regions for the *zen* sequences were identified by comparison to the *D. melanogaster* reference sequence. In our *D. melanogaster zen* sequences, the coding region was defined as positions 27–113 and 178–1,152. For *D. simulans zen,* the coding region was defined as positions 27–113 and 178–1,146.

Nucleotide diversity, $\pi$, is the average number of nucleotide differences per site between two randomly chosen sequences. It was calculated according to Nei [26] using DnaSP 4.10 [27]. To assess neutrality, we estimated Tajima's $D$ [17]. Then, using the observed Tajima's $D$ and the average number of nucleotide differences per sequence, we computed 95% confidence intervals by coalescent simulation in DnaSP.

## Supporting Information

**Table S1.** Stock Center Accession Numbers and Collection Locations of *Drosophila* Populations

Found at DOI: 10.1371/journal.pgen.0010057.st001 (41 KB DOC).

## References

1. Provine WB (1985) The R. A. Fisher – Sewall Wright controversy and its influence upon modern evolutionary biology. In: Dawkins R, Ridley M, editors. Oxford Surveys in Evolutionary Biology. Oxford: Oxford University Press. pp. 197–219
2. Seward AC, editor (1909) Darwin and modern science: Essays in the commemoration of the centenary of the birth of Charles Darwin and of the fiftieth anniversary of the publication of the Origin of Species. Cambridge: Cambridge University Press. 595 p.
3. Muller GB, Wagner GP (1991) Novelty in evolution—Restructuring the concept. Annu Rev Ecol Syst 22: 229–256.
4. Wagner GP, Muller GB (2002) Evolutionary innovations overcome ancestral constraints: A re-examination of character evolution in male sepsid flies (Diptera: Sepsidae). Evol Dev 4: 1–6.
5. Eberhard WG (2001) Multiple origins of a major novelty: Moveable abdominal lobes in male sepsid flies (Diptera: epsidae), and the question of developmental constraints. Evol Dev 3: 206–222.
6. Stauber M, Taubert H, Schmidt-Ott U (2000) Function of *bicoid* and *hunchback* homologs in the basal cyclorrhaphan fly *Megaselia* (Phoridae). Proc Natl Acad Sci U S A 97: 10844–10849.
7. Stauber M, Jackle H, Schmidt-Ott U (1999) The anterior determinant *bicoid*

of *Drosophila* is a derived *Hox* class 3 gene. Proc Natl Acad Sci U S A 96: 3786–3789.

8. Stauber M, Prell A, Schmidt-Ott U (2002) A single *Hox3* gene with composite *bicoid* and *zerknüllt* expression characteristics in non-Cyclorrhaphan flies. Proc Natl Acad Sci U S A 99: 274–279.

9. Falciani F, Hausdorf B, Schroeder R, Akam M, Tautz D, et al. (1996) Class 3 *Hox* genes in insects and the origin of *zen*. Proc Natl Acad Sci U S A 93: 8479–8484.

10. Hughes CL, Kaufman TC (2002) Exploring the myriapod body plan: Expression patterns of the ten *Hox* genes in a centipede. Development 129: 1225–1238.

11. Dearden P, Grbic M, Falciani F, Akam M (2000) Maternal expression and early zygotic regulation of the *Hox3*/*zen* gene in the grasshopper *Schistocerca gregaria*. Evol Dev 2: 261–270.

12. Schröder R, Sander K (1993) A comparison of transplantable *bicoid* activity and partial *bicoid* homeobox sequences in several *Drosophila* and blowfly species (Calliphoridae). Roux's Arch Dev Biol 203: 34–43.

13. Demuth JP, Wade M J (2005) Maternal expression increases the rate of *bicoid* evolution by relaxing selective constraint. Genetica. In press.

14. Whitlock MC, Wade MJ (1995) Speciation–Founder events and their effects on X-linked and autosomal genes. Am Nat 145: 676–685.

15. Wade MJ (2000) The evolutionary genetics of maternal effects. In: Mousseau TA, Fox CW, editors. Maternal effects as adaptations. Oxford: Oxford University Press. pp. 5–21.

16. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437–460.

17. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis. Genetics 123: 585–595.

18. Baines JF, Chen Y, Das A, Stephan W (2002) DNA sequence variation at a duplicated gene: Excess of replacement polymorphism and extensive haplotype structure in the *Drosophila melanogaster bicoid* region. Mol Biol Evol 19: 989–998.

19. Metz EC, Robles-Sikisaka R, Vacquier VD (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc Natl Acad Sci U S A 95: 10676–10681.

20. Wyckoff G, Wang W, Wu C-I (2000) Rapid evolution of male reproductive genes in the descent of man. Nature 403: 394–309.

21. Shuster SM, Wade MJ (2003) Mating systems and strategies. Princeton: Princeton University Press. 469 p.

22. Charlesworth B (1994) Evolution in age-structured populations. Cambridge: Cambridge University Press. 320 p.

23. Hughes KA, Alipaz JA, Drnevich JM, Reynolds RM (2002) A test of evolutionary theories of aging. Proc Natl Acad Sci U S A 99: 14286–14291.

24. Kalendar R (2005) FastPCR: A PCR primer design and repeat sequence searching software with additional tools for the manipulation and analysis of DNA and protein. Available at: http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm. Accessed: 04 November 2004.

25. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95–98.

26. Nei M (1987) Molecular evolutionary genetics. New York: Columbia University Press. 512 p.

27. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496–2497.